MASS SPECTROMETRY IN STRUCTURAL BIOLOGY AND BIOPHYSICS

WILEY SERIES ON MASS SPECTROMETRY

Series Editors

Dominic M. Desiderio
Departments of Neurology and Biochemistry
University of Tennessee Health Science Center

Nico M. M. Nibbering
Vrije Universiteit Amsterdam, The Netherlands

A complete list of the titles in this series appears at the end of this volume.

MASS SPECTROMETRY IN STRUCTURAL BIOLOGY AND BIOPHYSICS

Architecture, Dynamics, and Interaction of Biomolecules

Second Edition

IGOR A. KALTASHOV
STEPHEN J. EYLES
University of Massachusetts at Amherst

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Kaltashov, Igor A.
Mass spectrometry in structural biology and biophysics : architecture, dynamics, and interaction of biomolecules / Igor A Kaltashov, Stephen J Eyles. – 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-93779-2 (cloth)
1. Mass spectrometry. 2. Biophysics. 3. Biomolecules–Spectra. I. Eyles, Stephen J. II. Title.
QP519.9.M3K35 2012
610.28'4–dc23

2011040425

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTENTS

PREFACE TO THE FIRST EDITION

Strictly speaking, the term *biophysics* refers to the application of the theories and methods of physics to answer questions in the biological arena. This obviously now vast field began with studies of how electrical impulses are transmitted in biological systems and how the shapes of biomolecules enable them to perform complex biological functions. Over time, biophysicists have added a wide variety of methodologies to their experimental toolkit, one of the more recent additions being mass spectrometry (MS). Traditionally limited to the analysis of small molecules, recent technological advances have enabled the field of MS to expand into the biophysical laboratory, catalyzed by the 2002 Nobel prize winning work of John Fenn and Koichi Tanaka. Mass spectrometry is a rapidly developing field whose applications are constantly changing. This text represents only a snapshot of current techniques and methodologies.

This book aims to present a detailed and systematic coverage of the current state of biophysical MS with special emphasis on experimental techniques that are used to study protein higher order structure and dynamics. No longer an exotic novelty, various MS based methods are rapidly gaining acceptance in the biophysical community as powerful experimental tools to probe various aspects of biomolecular behavior both *in vitro* and *in vivo*. Although this field is now experiencing an explosive growth, there is no single text that focuses solely on applications of MS in molecular biophysics and provides a thorough summary of the plethora of MS experimental techniques and strategies that can be used to address a wide variety of problems related to biomolecular dynamics and higher order structure. This book aims to close that gap.

We intended to target two distinct audiences: mass spectrometrists who are working in various fields of life sciences (but are not necessarily experts in biophysics) and experimental biophysicists (who are less familiar with recent developments in MS technology, but would like to add it to their experimental arsenal). In order to make the book equally useful for both groups, the presentation of the MS based techniques in biophysics is preceded by a discussion of general biophysical concepts related to the structure and dynamics of biological macromolecules (Chapter 1). Although it is not meant to provide an exhaustive coverage of the entire field of molecular biophysics, the fundamental concepts are explained in some detail to enable anyone not directly involved with the field to understand the important aspects and terminology. Chapter 2 provides a brief overview of "traditional" biophysical techniques with special emphasis on those that are complementary to MS and that are mentioned elsewhere in the book. These introductory chapters are followed by an in-depth discussion of modern mass spectrometric hardware used in experimental studies of biomolecular structure and dynamics. The purpose of Chapter 3 is to provide readers who are less familiar with MS with concise background material on modern MS instrumentation and techniques that will be referred to in the later chapters (the book is structured in such a way that no prior familiarity with biological MS is required of the reader).

Chapters 4–7 deal with various aspects of protein higher order structure and dynamics as probed by various MS based methods. Chapter 4 focuses on "static structures", by considering various approaches to evaluate higher order structure of proteins at various levels of spatial resolution when crystallographic and nuclear magnetic resonance (NMR) data are either unavailable or insufficient. The major emphasis is on methods that are used to probe biomolecular topology and solvent accessibility (i.e., chemical cross-linking and

selective chemical modification). In addition, the use of hydrogen–deuterium exchange for mapping protein–protein interfaces is briefly discussed. Chapter 5 presents a concise introduction to an array of techniques that are used to study structure and behavior of non-native protein states that become populated under denaturing conditions. The chapter begins with consideration of protein ion charge state distributions in electrospray ionization mass spectra as indicators of protein unfolding and concludes with a detailed discussion of hydrogen exchange, arguably one of the most widely used methods to probe the structure and dynamics of non-native protein states under equilibrium conditions. The kinetic aspects of protein folding and enzyme catalysis are considered in Chapter 6. Chapter 7 focuses on MS based methods that are used to extract quantitative information on protein–ligand interactions (i.e., indirect methods of assessment of binding energy). The remainder of this chapter is devoted to advanced uses of MS to characterize dynamics of multiprotein assemblies and its role in modulating protein function.

Complementarity of MS based techniques to other experimental tools is emphasized throughout the book and is also addressed specifically in Chapter 8. Two examples presented in this chapter are considered in sufficient detail to illustrate the power of synergy of multiple biophysical techniques, where some methods provide overlapping information to confirm the evidence, while others provide completely unique details. Chapter 9 presents a discussion of MS based methods to study the higher order structure and dynamics of biopolymers that are not proteins (oligonucleotides, polysaccharides, as well as polymers of nonbiotic origin). Chapter 10 provides a brief discussion of biomolecular properties in the gas phase, focusing primarily on the relevance of *in vacuo* measurements to biomolecular properties in solution.

This book concludes with a discussion of the current challenges facing biomolecular MS, as well as important new developments in the field that are not yet ready for routine use. Chapter 11 focuses on several areas where MS is currently making a debut. It begins with a discussion of novel uses of MS aimed at understanding "orderly" protein oligomerization processes, followed by consideration of "catastrophic" oligomerization (e.g., amyloidosis). This chapter also considers other challenging tasks facing modern MS, such as the detection and characterization of very large macromolecular assemblies (e.g., intact ribosomes and viral particles), as well as applications of various MS based techniques to study the behavior of a notoriously difficult class of biopolymers–membrane proteins. This chapter concludes with a general discussion of the relevance of *in vitro* studies and reductionist models to processes occurring *in vivo*.

Throughout the entire book, an effort has been made to present the material in a systematic fashion. Both the theoretical background and technical aspects of each technique are discussed in detail, followed by an outline of its advantages and limitations, so that the reader can get a clear sense of both current capabilities and potential future uses of various MS based experimental methodologies. Furthermore, this book was conceived as a combination of a textbook, a good reference source, and a practical guide. With that in mind, a large amount of material (practical information) has been included throughout. An effort has also been made to provide the reader with a large reference base to original research papers, so that the details of experimental work omitted in the book can easily be found. Because of space limitations and the vastness of the field, a significant volume of very interesting and important research could not be physically cited. It is hoped, however, that no important experimental techniques and methodologies have been overlooked. The authors will be grateful for any comments from the readers on the material presented in the book (Chapters 1, 3, 4, 5, 7, 10, and 11 were written mostly by I.K. and Chapters 6, 8, and 9 by S.E.; both authors contributed equally to Chapter 2). The comments can be e-mailed directly to the authors at kaltashov@chem.umass.edu and eyles@polysci.umass.edu.

We are grateful to Professors David L. Smith, Michael L. Gross, Max Deinzer, Lars Konermann, Joseph A. Loo, and Richard W. Vachet for helpful discussions over the past several years that have had direct impact on this book. We would also like to thank many other colleagues, collaborators, and friends for their support and encouragement during various stages of this challenging project. We are also indebted to many people who have made contributions to this book in the form of original graphics from research articles (the credits are given in the relevant parts of the text). We also thank the current and past members of our research group, who in many cases contributed original unpublished data for the illustrative material presented throughout. Finally, we would like to acknowledge the National Institutes of Health and the National Science Foundation for their generous support of our own research efforts at the interface of biophysics and mass spectrometry.

IGOR A. KALTASHOV

STEPHEN J. EYLES

University of Massachusetts at Amherst

PREFACE TO THE SECOND EDITION

The first edition of *Mass Spectrometry in Biophysics* was published over six years ago, and this field has experienced a truly transformative change during this period. Investigation of architecture and behavior of biopolymers (mostly proteins) by mass spectrometry (MS) was performed in the early 2000s in only a handful of laboratories around the world. The results of these studies were frequently met with skepticism outside of the MS community. However, by the end of the decade, MS had become an indispensable tool in experimental biophysics, which is capable of providing unique information on the conformation and dynamics of biopolymers, as well as their interactions with physiological partners. Not only has MS continued to progress at an accelerated pace throughout these years, but the scope of its applications in biophysics and structural biology also expanded very dramatically. As a result of these developments, some of the segments of the first edition became somewhat outdated and no longer provided adequate coverage of several key state-of-the-art techniques.

Another exciting change that has occurred in recent years is that MS-based studies of protein behavior are no longer confined to the realm of academic science. Indeed, the explosive growth of the biopharmaceutical sector in the past decade brings to the fore the need to have capabilities to analyze behavior of protein therapeutics and places a premium on developing analytical techniques able to handle these extremely complex species. Mass spectrometry can certainly fit the bill, and the gradual acceptance of these new tools within the biopharmaceutical industry and regulatory agencies as reliable methods to study architecture and dynamics of biomolecules is of little surprise to anyone.

In preparing the second edition of this book, our aim was to bring the reader up to date with the field by providing an expanded and up-to-date coverage of MS-based experimental methodologies in biophysics and structural biology, as well as addressing the specific needs of the new and rapidly growing segment of practitioners of this technique in the biopharmaceutical industry. We have tried as much as possible to preserve the original organization of the book, which proved very efficient in presenting the material. Introductory Chapters 1 and 2 were minimally changed, while Chapter 3 was updated to reflect, *inter alia*, introduction and rapid proliferation of Orbitrap mass analyzers and ion mobility spectrometers, as well as wide acceptance of the so-called electron-based ion fragmentation techniques (e.g., electron capture and electron transfer dissociation). The most extensive changes were made to Chapters 4–7, which present experimental methodologies used to probe various aspects of protein architecture and behavior under a variety of conditions. Similarly, very extensive revision was made to Chapter 8 (Chapter 9 in the first edition), which reflects a continuing expansion of MS into the realm of oligonucleotides, polysaccharides, and synthetic polymers, as well as polymer–protein conjugates.

Former Chapters 8 and 10 from the first edition were removed from this book. Indeed, the synergism between MS and other biophysical techniques (the topic of the former Chapter 8) is now commonly accepted, and in fact has become a defining element in the experimental design; many examples of this are dispersed throughout the text of the second edition. Studies of structure and behavior of biopolymers in the gas phase (the topic of the former Chapter 10) have now transformed into a separate field, and its careful and detailed consideration is no longer possible in this book given obvious space limitations. The exception is made for several gas-phase methods that are either already used to study solution structure (e.g., gas-phase H/D exchange to probe oligonucleotide conformations) or show promise in that

regard (e.g., measurements of biopolymer ion mobility in the gas phase). The final chapter of this book (chapter 9, which was old chapter 11) once again strives to go beyond routine measurements and considers several fields that are currently out of the reach of the commonly accepted MS based techniques (membrane proteins, protein aggregates, very large biopolymer assemblies, etc.).

Taken together, the second edition is a systematic presentation of a modern mass spectrometry-based armamentarium that can be used to solve a variety of challenging problems in biophysics, structural biology, and biopharmaceuticals. One of our goals was not only to provide practical advice, but also to arm the reader with a solid coverage of all relevant fundamental issues, including extensive references to, and examples from, the original published work. In addition to that, the book contains a large number of examples and illustrations taken from the work carried out in our laboratory, some of which have never been published. We are indebted to the past and present group members who provided this material (names in parentheses indicate present employment if different from UMass-Amherst): Dr. Dmitry R. Gumerov (Mersana Pharmaceuticals), Dr. Andras Dobo (Sigma-Aldrich-Fluka Europe), Prof. Hui Xiao (Albert Einstein School of Medicine), Dr. Anirban Mohimen (Vertex Pharmaceuticals), Prof. Wendell Griffith (University of Toledo), Dr. Joshua K. Hoerner (Schering-Plough Research Institute), Dr. Mingxuan "Sunshine" Zhang (Biogen IDEC), Dr. Virginie Sjoelund (National Institutes of Health), Dr. Rachael Leverence (University of Wisconsin), Dr. Agya Frimpong, Dr. Rinat R. Abzalimov, Dr. Cedric E. Bobst, Mr. Guanbo Wang, and Mr. Shunhai Wang.

We are also grateful to Professors Michael L. Gross (Washington University at St. Louis), S. Walter Englander (University of Pennsylvania Medical School), George H. Lorimer (University of Maryland at College Park), Virgil L. Woods, Jr. (University of California at San Diego School of Medicine), Roman A. Zubarev (Karolinska Institute), Lars Konermann (University of Western Ontario), Joseph A. Loo (UCLA), John Engen (Northeastern University), and Richard W. Vachet (UMass-Amherst) for numerous very helpful discussions over the past several years that have had direct impact on this book. We would also like to acknowledge our collaborators from industry who helped us better understand the unique needs of the biopharmaceutical sector and how they can be addressed using mass spectrometry tools: Dr. Pavel Bondarenko (Amgen, Inc.), Drs. Steven Berkowitz and Damian Houde (Biogen IDEC), and Drs. Philip Savickas, John Thomas, Melanie Lin, and Paul Salinas (Shire Human Genetic Therapies). Finally, we would like to acknowledge the National Institutes of Health and the National Science Foundation for their generous support of our own research efforts at the interface of biophysics and MS, many examples of which are presented in this book.

IGOR A. KALTASHOV

STEPHEN J. EYLES

University of Massachusetts at Amherst

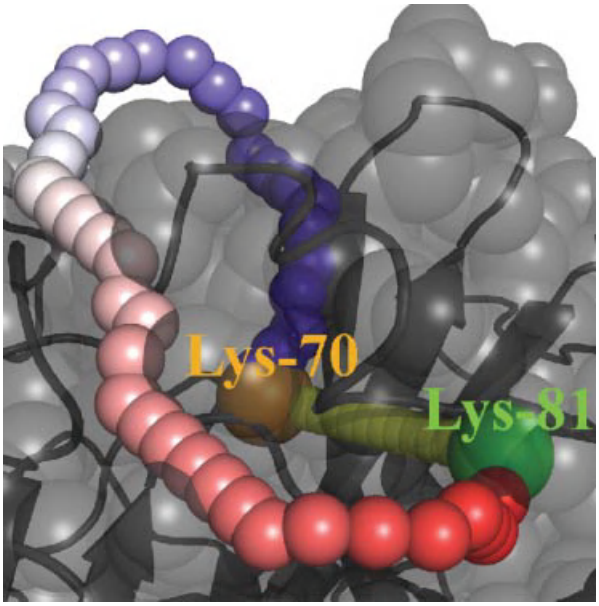


Figure 4.11 Shortest solvent accessible surface distance (SASD) path illustrated with human prothrombin. (*See text for full caption.*)

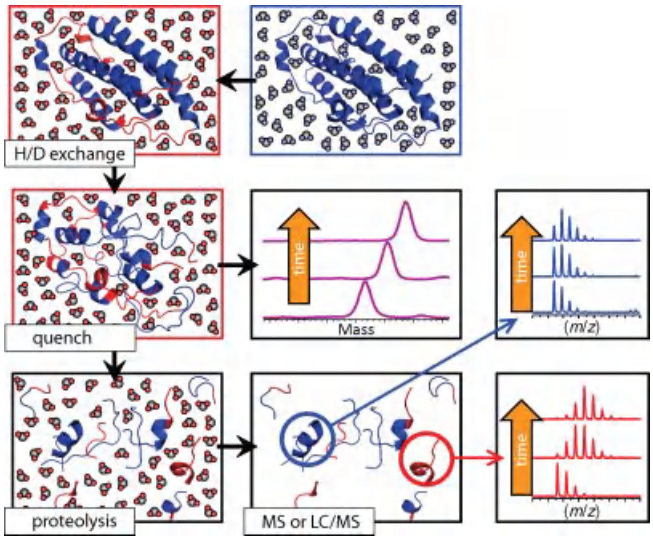


Figure 4.16 Schematic representation of HDX MS work flow to examine protein higher order structure and conformational dynamics. (*See text for full caption.*)

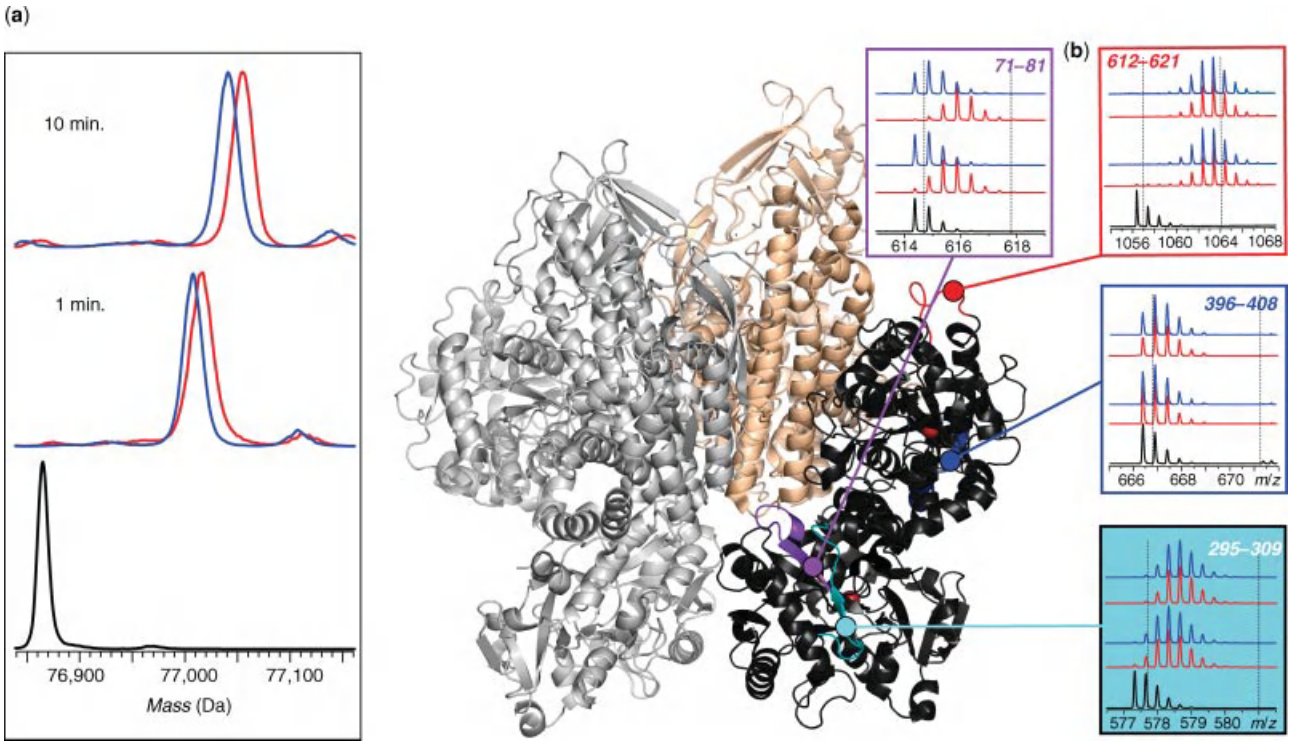


Figure 4.17 Localization of the receptor binding interface on the surface of human serum transferrin (Tf) with HDX MS. (a) The HDX MS of Tf (global exchange) in the presence (blue) and the absence (red) of the receptor. (*See text for full caption.*)

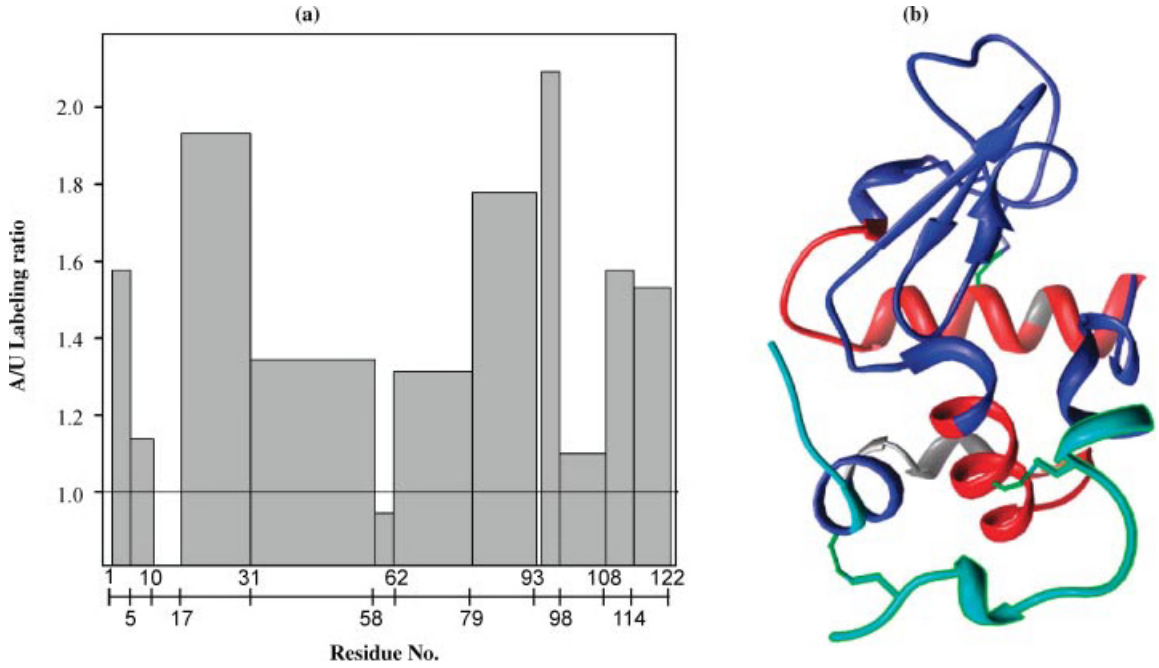


Figure 5.7 (*a*) Regiospecific :CH_2 labeling pattern of the acid-stabilized molten globule (A-state) of bovine α-lactalbumin relative to that of the unfolded protein (U-state). (*See text for full caption.*)

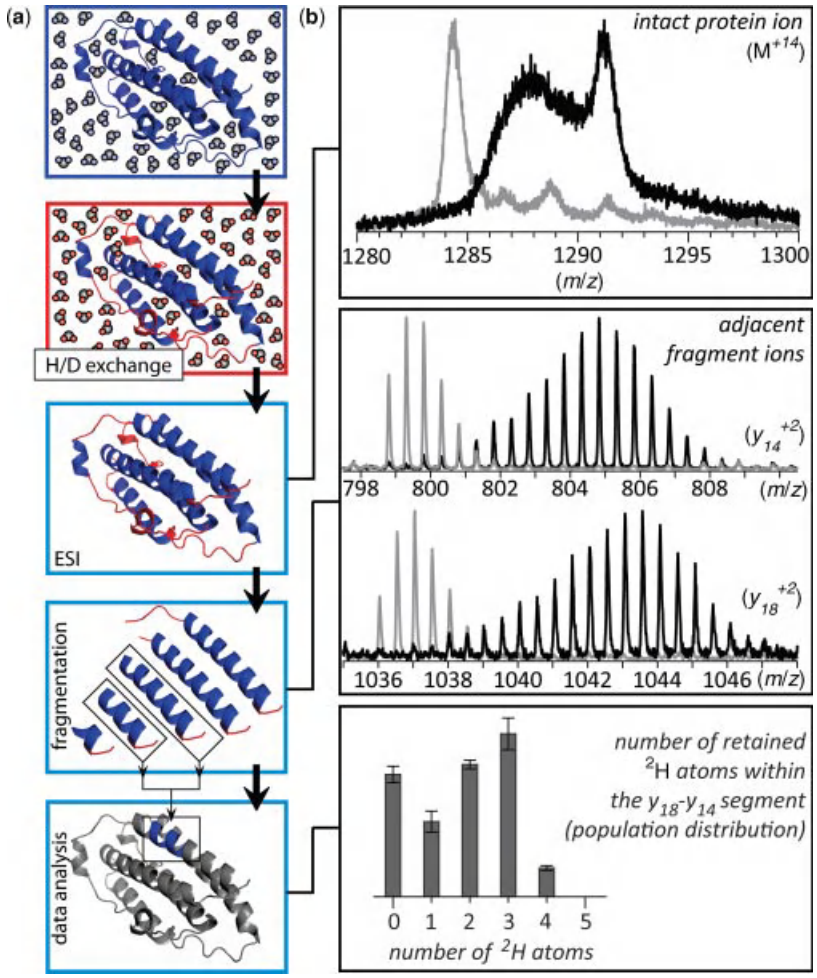


Figure 5.22 A flow-chart diagram of the top-down HDX MS (HDX MS/MS) measurements (*a*) and the steps involved in data processing (*b*). The data represent HDX MS/MS measurements of *pseudo-wild-type(wt*)* cellular retinoic acid binding protein I under mildly denaturing conditions. The existence of multiple protein conformers is evident from the convoluted appearance of the isotopic distribution of the intact protein ion). Protection in the C-terminal protein segments can be deduced from the isotopic distributions of corresponding *y*-fragments (*See text for full caption.*)

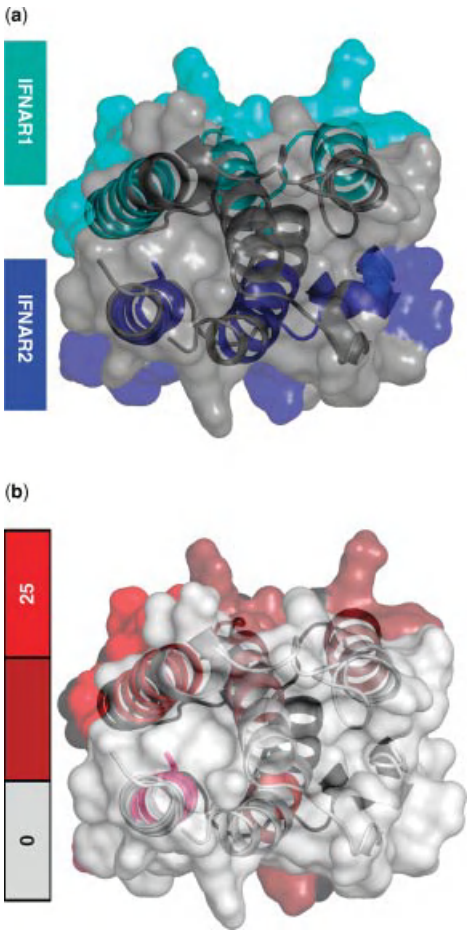


Figure 7.13 (*a*) Receptor binding interfaces of IFN from earlier mutagenesis work (123). (*b*) Regions affected by NEM alkylation. (*See text for full caption.*)

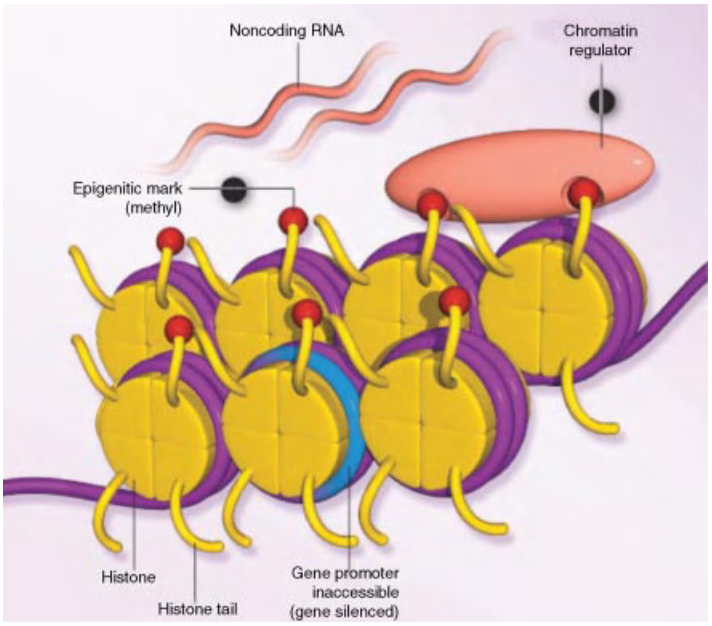


Figure 8.2 A schematic representation of The DNA packaging in chromatin. (*See text for full caption.*)

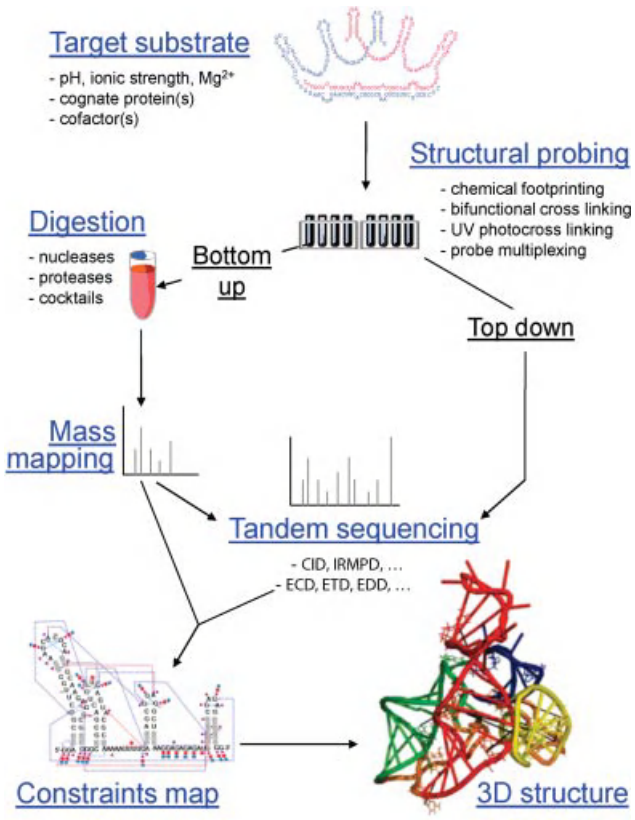


Figure 8.8 General workflow for three-dimensional (3D)-structure determination of nucleic acids based on structural probing and MS analysis (MS3D). (*See text for full caption.*)

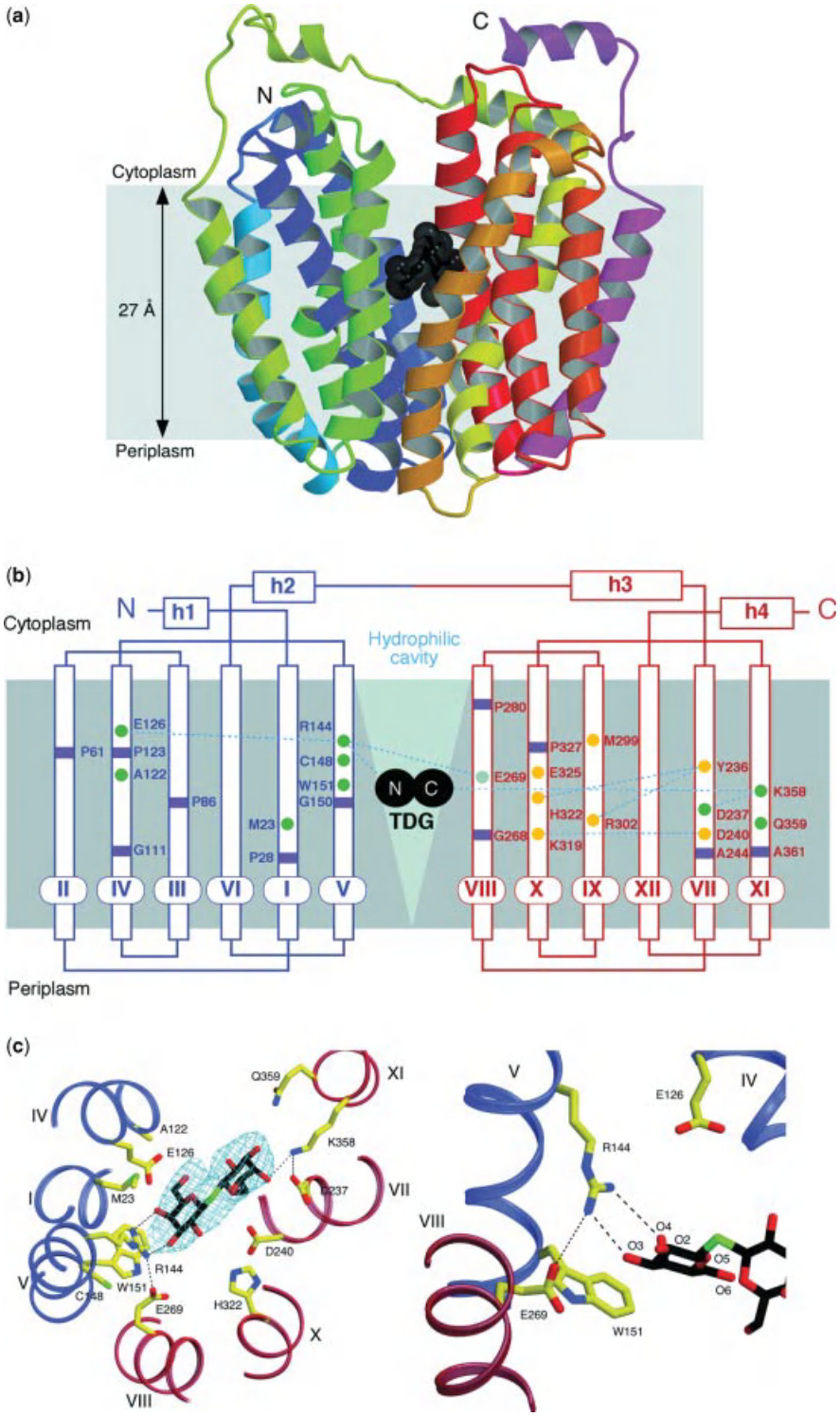


Figure 9.4 Overall structure of lactose permease from *E. coli* (LacY) with a bound substrate homologue TDG. (*a*) Ribbon representation of LacY viewed parallel to the membrane. (*b*) Secondary structure schematic. (*c*) Substrate binding site. (*See text for full caption.*)

1

GENERAL OVERVIEW OF BASIC CONCEPTS IN MOLECULAR BIOPHYSICS

This introductory chapter provides a brief overview of the basic concepts and current questions facing biophysicists in terms of the structural characterization of proteins, protein folding, and protein–ligand interactions. Although this chapter is not meant to provide an exhaustive coverage of the entire field of molecular biophysics, the fundamental concepts are explained in some detail to enable anyone not directly involved with the field to understand the important aspects and terminology.

1.1. COVALENT STRUCTURE OF BIOPOLYMERS

Biopolymers are a class of polymeric materials that are manufactured in nature. Depending on the building blocks (or *repeat units* using polymer terminology), biopolymers are usually divided into three large classes. These are (1) polynucleotides (built of nucleotides); (2) peptides and proteins (built of amino acids); and (3) polysaccharides (built of various saccharide units). This chapter only considers general properties of biopolymers using peptides and proteins as examples; questions related to polynucleotides and polysaccharides will be discussed in some detail in Chapter 8.

All polypeptides are linear chains built of small organic molecules called *amino acids*. There are 20 amino acids that are commonly considered *canonical* or *natural* (Table 1.1). This assignment is based upon the fact that these 20 amino acids correspond to 61 (out of total 64) codons within the triplet genetic code with three remaining codons functioning as terminators of protein synthesis (1,2), although there are at least as many other amino acids that occur less frequently in living organisms (Table 1.2). Noncanonical amino acids are usually produced by chemical modification of a related canonical amino acid (e.g., oxidation of proline produces hydroxyproline), although at least two of them (selenocysteine and pyrrolysine) should be considered canonical based on the way they are utilized in protein synthesis *in vivo* by some organisms (3,4). Furthermore, new components can be added to the protein biosynthetic machinery of both prokaryotes and eukaryotes, which makes it possible to genetically encode unnatural amino acids *in vivo* (5,6). A peculiar structural feature of all canonical (with the exception of glycine) and most noncanonical amino acids is the presence of an asymmetric carbon atom (C_α), which should give rise to two different enantiomeric forms. Remarkably, all canonical amino acids are of the L-type. The D-forms of amino acids can also be synthesized *in vivo*, and are particularly abundant in fungi; however, these amino acids do not have access to the genetic code. The rise and persistence of homochirality in the living world throughout the entire evolution of life remains one of the greatest puzzles in biology; examples of homochirality at the molecular level also include almost exclusive occurrence of the D-forms of sugars in the nucleotides, while manifestations of homochirality at the macroscopic level range from specific helical patterns of snail shells to the chewing motions of cows (7,8).

Unlike most synthetic polymers and structural biopolymers (several examples of which will be presented in Chapter 8), peptides and proteins have a very specific sequence of monomer units. Therefore, even though polypeptides can be considered simply as highly functionalized linear polymers constituting a nylon-2 backbone, these functional groups, or *side chains*, are arranged in a highly specific order. All

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

TABLE 1.1. Chemical Structure and Masses of Natural (Canonical) Amino Acids

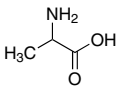
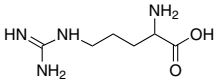
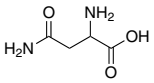
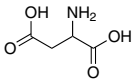
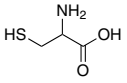
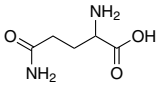
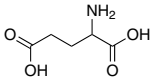
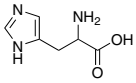
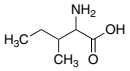
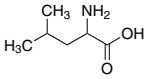
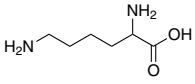
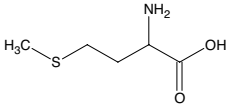
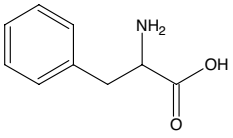
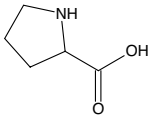
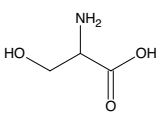
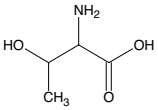
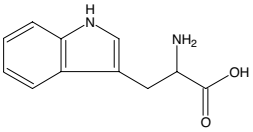
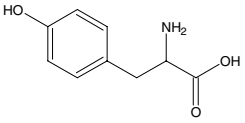
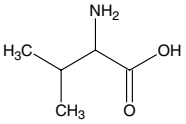
Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass[a] (Residue)	Average Mass (Residue)
Ala (A)	Alanine	C_3H_5NO		Nonpolar	71.037	71.079
Arg (R)	Arginine	$C_6H_{12}N_4O$		Basic	156.101	156.188
Asn (N)	Asparagine	$C_4H_6N_2O_2$		Polar	114.043	114.104
Asp (D)	Aspartic acid	$C_4H_5NO_3$		Acidic	115.027	115.089
Cys (C)	Cysteine	C_3H_5NOS		Polar/acidic	103.009	103.145
Gln (Q)	Glutamine	$C_5H_8N_2O_2$		Polar	128.059	128.131
Glu (E)	Glutamic acid	$C_5H_7NO_3$		Acidic	129.043	129.116
Gly (G)	Glycine	C_2H_3NO		Nonpolar	57.021	57.052
His (H)	Histidine	$C_6H_7N_3O$		Basic	137.059	137.141
Ile (I)	Isoleucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Leu (L)	Leucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Lys (K)	Lysine	$C_6H_{12}N_2O$		Basic	128.095	128.174
Met (M)	Methionine	C_5H_9NOS		Nonpolar/ amphipathic	131.040	131.199
Phe (F)	Phenylalanine	C_9H_9NO		Nonpolar	147.068	147.177
Pro (P)	Proline	C_5H_7NO		Nonpolar	97.053	97.117
Ser (S)	Serine	$C_3H_5NO_2$		Polar	87.032	87.078

TABLE 1.1. (*Continued*)

Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass[a] (Residue)	Average Mass (Residue)
Thr (T)	Threonine	$C_4H_7NO_2$		Polar/amphipathic	101.048	101.105
Trp (W)	Tryptophan	$C_{11}H_{10}N_2O$		Amphipathic	186.079	186.213
Tyr (Y)	Tyrosine	$C_9H_9NO_2$		Amphipathic	163.063	163.176
Val (V)	Valine	C_5H_9NO		Nonpolar	99.068	99.133

[a] See Chapter 3 for a definition of monoisotopic and average masses.

naturally occurring proteins consist of an exact sequence of amino acid residues linked by peptide bonds (Fig. 1.1*a*), which is usually referred to as the *primary structure*. Some amino acids can be modified after translation (termed *post-translational modification*), for instance, by phosphorylation, methylation, or glycosylation. Among these modifications, formation of the covalent bonds between two cysteine residues is particularly interesting, since such *disulfide bridges* can stabilize protein geometry, by bringing together residues that are distant in the primary structure into close proximity in three-dimensional (3D) space. The highly specific spatial organization of many (but not all) proteins under certain conditions is often referred to as *higher order structure* and is another point of distinction between them (as well as most biological macromolecules) and synthetic polymers. Although disulfide bridges are often important contributors to the stability of the higher order structure, correct protein folding does not necessarily require such covalent "stitches". In fact, cysteine is one of the least abundant amino acids, and many proteins lack it altogether. As it turns out, relatively weak noncovalent interactions between functional groups of the amino acid side chains and the polypeptide backbone are much more important for the highly specific arrangement of the protein in 3D space. Section 1.2 provides a brief overview of such interactions.

1.2. NONCOVALENT INTERACTIONS AND HIGHER ORDER STRUCTURE

Just like all chemical forces, all inter- and intramolecular interactions involving biological macromolecules (both covalent and noncovalent) are electrical in nature and can be described generally by the superposition of Coulombic potentials. In practice, however, the noncovalent interactions are subdivided into several categories, each being characterized by a set of unique features.

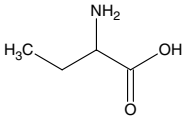
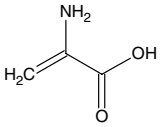
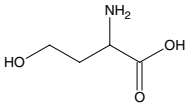
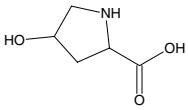
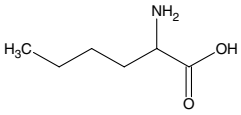
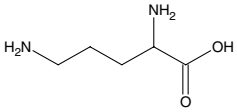
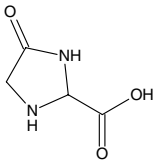
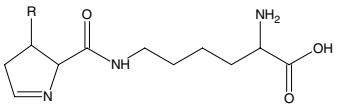
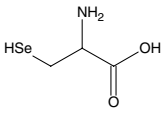
1.2.1. Electrostatic Interaction

The term *electrostatic interaction* broadly refers to a range of forces exerted among a set of stationary charges and/or dipoles. The interaction between two fixed charges q_1 and q_2 separated by a distance r is given by the Coulomb law:

$$E = \frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r} \tag{1-2-1}$$

where ε_0 is the absolute permittivity of vacuum [$8.85 \times 10^{-12}\,C^2/N{\cdot}m$ in Système International (SI)] and ε is the dielectric constant of the medium. Although the numerical values of the dielectric constants of most homogeneous media are readily available, the use of this concept at the microscopic level is not very straightforward (9,10). The dielectric constant is a measure of the screening of the electrostatic interaction due to the polarization of the medium, hence the difficulty in defining a single constant for a protein, where such screening depends on the exact location of the charges, their environment, and so on. Although in some cases the values of the "effective" dielectric constants for specific protein systems can be estimated based on experimental measurements of the electrostatic interactions, such an approach has been disfavored by many for a long time (11). This book will follow the example set by Daune (12) and will write all expressions with $\varepsilon = 1$.

TABLE 1.2. Chemical Structure and Masses of Some Less Frequently Occurring Natural (Noncanonical) Amino Acids

Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass (Residue)	Average Mass (Residue)
Abu	2-Aminobutyric acid	C_4H_7NO		Nonpolar	85.053	85.106
Dha	Dehydroalanine	C_3H_3NO		Nonpolar	69.021	69.063
Hse	Homoserine	$C_4H_7NO_2$		Polar	101.048	101.105
Hyp	Hydroxyproline	$C_6H_{12}N_2O_2$		Polar	144.090	144.174
Nle	Norleucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Orn	Ornithine	$C_5H_{10}N_2O$		Basic	114.079	114.147
Pyr	Pyroglutamic acid	$C_5H_5NO_2$		Moderately polar	111.032	111.100
Pyl	Pyrrolysine	$C_{11}H_{16}N_3O_2 + R$ (NH_2, OH, or CH_3)		Polar		
Sec	Selenocysteine	C_3H_5NOSe		Polar/acidic	144.960 (150.954[a])	150.039

[a] Most abundant.

Interaction between a charge q and a permanent dipole p separated by a distance r is given by

$$E = -\frac{qp \cdot \cos\theta}{4\pi\varepsilon_0 r^2} \qquad (1\text{-}2\text{-}2)$$

where θ is the angle between the direction of the dipole and the vector connecting it with the charge q. If the dipole is not fixed directionally, it will align itself to minimize the energy Eq. (1-2-2), that is, $\theta = 0$. However, if such energy is small compared to thermal energy, Brownian motion will result in

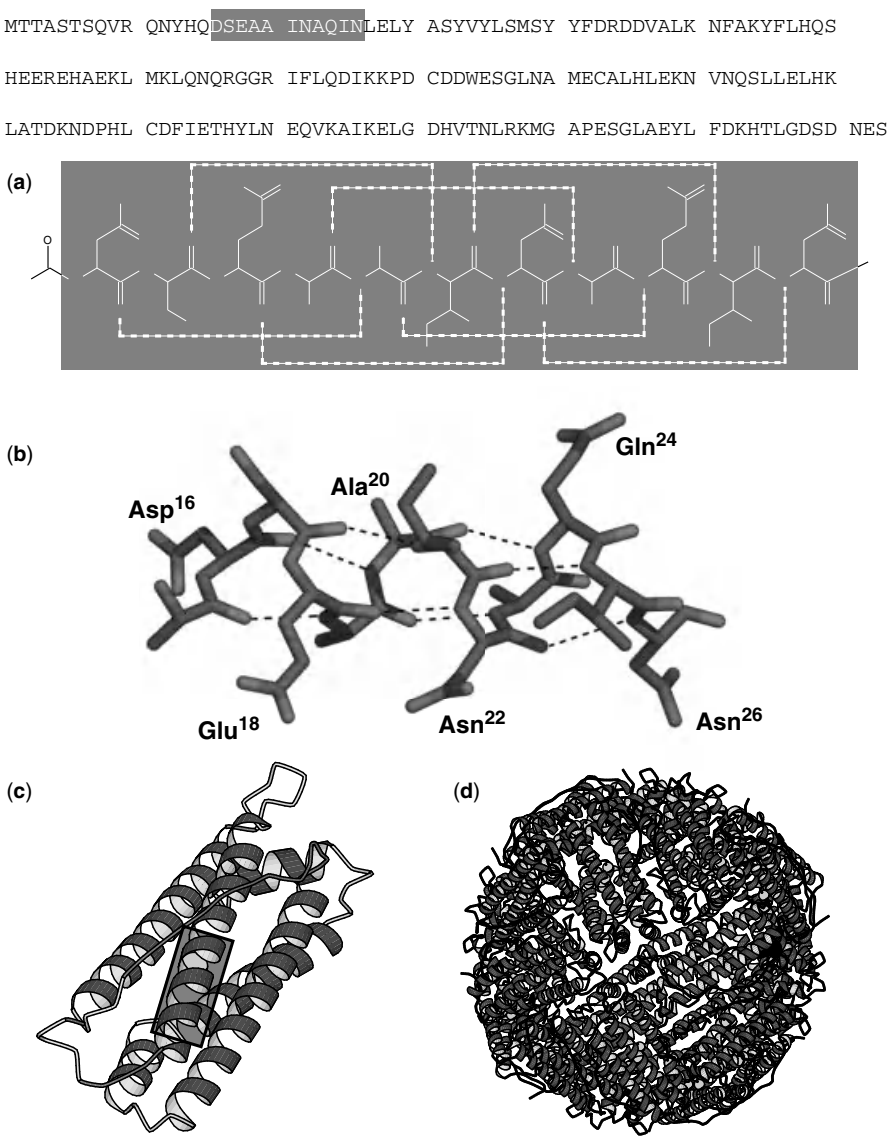


Figure 1.1. Hierarchy of structural organization of a protein (H-form of human ferritin). Amino acid sequence determines the primary structure (*a*). Covalent structure of the 11 amino acid residue long segment of the protein ($Glu^{16} \rightarrow Asn^{26}$) is shown in the shaded box. A highly organized network of hydrogen bonds along the polypeptide backbone (shown with dotted lines) gives rise to secondary structure, α-helix (*b*). A unique spatial arrangement of the elements of the secondary structure gives rise to the tertiary structure, with the shaded box indicating the position of the ($Glu^{16} \rightarrow Asn^{26}$) segment (*c*). Specific association of several folded polypeptide chains (24 in the case of ferritin) produces the quaternary structure (*d*).

the averaging of all values of θ with only a small preference for those that minimize the electrostatic energy, resulting in a much weaker overall interaction:

$$E = -\frac{q^2 p^2}{(4\pi\varepsilon_0)^2 \cdot 3k_B T r^4} \qquad (1\text{-}2\text{-}3)$$

where $T =$ temperature and k_B is the Boltzmann constant.

Interaction between two dipoles, p_1 and p_2, separated by a distance r in this approximation will be given by

$$E = -\frac{2p_1^2 p_2^2}{(4\pi\varepsilon_0)^2 \cdot 3k_B T r^6} \qquad (1\text{-}2\text{-}4)$$

while the interaction between the two fixed dipoles will be significantly stronger ($\sim 1/r^3$).

Polarization of a molecule can also be viewed in terms of electrostatic interaction using a concept of *induced dipoles* (12). Such interaction is, of course, always an attractive force, which is inversely proportional to r^4 (for a charge-induced dipole interaction) or r^6 (for a permanent dipole–induced dipole interaction). Finally, interaction between two polarizable molecules can be described in terms of a weak induced dipole–induced dipole interaction.

1.2.2. Hydrogen Bonding

The electrostatic interactions considered in the preceding sections can be treated using classical physics. *Hydrogen bonding* is an example of a specific noncovalent interaction that cannot be treated within the framework of classical electrostatics. It refers to an interaction occurring between a proton donor group ($-$OH, $-NH_3^+$, etc.) and a proton acceptor atom that has an unshared pair of electrons. Although hydrogen-bond formation (e.g., $R=\ddot{O}:\cdots H-NR_2$) may look like a simple electrostatic attraction of the permanent dipole–induced dipole type, the actual interaction is more complex and involves charge transfer within the proton donor–acceptor complex. The accurate description of such exchange interaction requires the use of sophisticated apparatus of quantum mechanics.

The importance of hydrogen bonding as a major determinant and a stabilizing factor for the higher order structure of proteins was recognized nearly 70 years ago by Mirsky and Pauling, who wrote in 1936: "*the* [native protein] *molecule consists of one polypeptide chain which continues without interruption throughout the molecule . . . this chain is folded into a uniquely defined configuration, in which it is held by hydrogen bonds between the peptide nitrogen and oxygen atoms* . . ." (13). Considerations of the spatial arrangements that maximize the amount of hydrogen bonding within a polypeptide chain later led Pauling to predict the existence of the α-helix, one of the most commonly occurring local motifs of higher order structure in proteins (14). Hydrogen bonds can be formed not only within the macromolecule itself, but also between biopolymers and water molecules (the latter act as both proton donors and acceptors). Hydrogen bonding is also central for understanding the physical properties of water, as well as other protic solvents.

1.2.3. Steric Clashes and Allowed Conformations of the Peptide Backbone: Secondary Structure

Both electrostatic and hydrogen-bonding interactions within a flexible macromolecule would favor 3D arrangements of its atoms that minimize the overall potential energy. However, there are two fundamental restrictions that limit the conformational freedom of the macromolecule. One is, of course, the limitation imposed by covalent bonding. The second is steric hindrance, which also restricts the volume of conformational space available to the biopolymer. This section considers the limits imposed by steric clashes on the conformational freedom of the polypeptide backbone.

The peptide amide bond is represented in Figure 1.1*a* as a single bond (i.e., C$-$N), however, it actually has a partial double-bond character in a polypeptide chain due to partial delocalization of electron density across the neighboring carbonyl group. The double-bond character of the C$-$N linkage, as well as the strong preference for the trans configuration of the amide hydrogen and carbonyl oxygen atoms,* result in four atoms lying coplanar. Figure 1.2 shows successive planes linked by the C_α atom of the *i*th amino acid residue. The two degrees of freedom at this junction are usually referred to as φ_i and ψ_i angles and the backbone conformation of the polypeptide composed of n amino acid residues can be described using $n-1$ parameters (pairs of φ_i and ψ_i). Steric restrictions limit the conformational volume accessible to polypeptides, which is usually represented graphically on the (φ, ψ) plane using so-called *conformational maps* or *Ramachandran plots* (15). An example of such a diagram, shown in Figure 1.3, clearly indicates that only a very limited number of configurations of the polypeptide backbone are allowed sterically.

Several regions within the accessible conformational volume are of particular interest, since they represent the structures that are stabilized by highly organized networks of hydrogen bonds. The α-helix is one of such structures, where the carbonyl oxygen atom of the *i*th residue is hydrogen bonded to the amide of the $(i+4)$th residue (Fig. 1.1*b*). This local motif, or spatial arrangement of a segment of the polypeptide backbone, is an example of a *secondary structure*, which is considered the first stage of macromolecular organization to form-higher order structure. Another commonly occurring element of the secondary structure is located within a larger island of sterically allowed conformations on the Ramachandran plot. Such conformations [upper left corner on the (φ, ψ) plane in Fig. 1.3] are rather close to the fully extended configuration of the chain and, therefore, cannot be stabilized by local hydrogen bonds. Nevertheless, formation of strong stabilizing networks of hydrogen bonds becomes possible if two strands are placed parallel or antiparallel to each other, forming so-called *β-pleated sheets*.

*Proline is an exception to this rule. As an imino acid its side chain is also bonded to the nitrogen atom. Thus, the cis and trans forms are almost isoenergetic, leading to the possibility of *cis*-Xaa$-$Pro bonds in folded proteins, and statistically at the level of 5–30% in unstructured polypeptides.

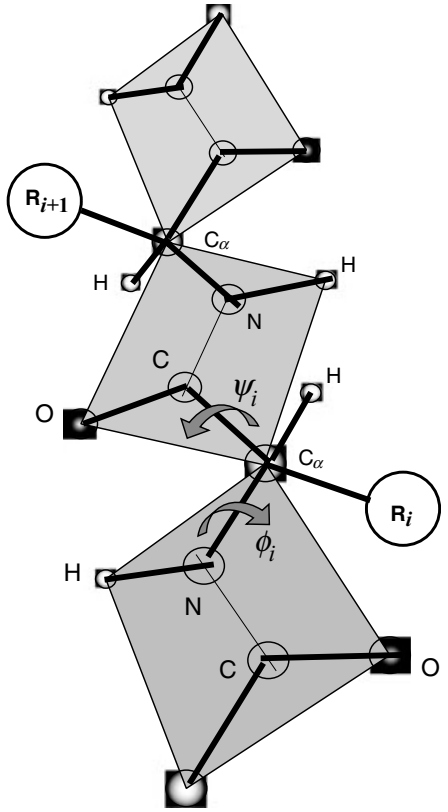


Figure 1.2. Peptide bond and the degrees of freedom determining the polypeptide backbone conformation.

The third important local structural motif is the *turn*, which causes a change in the chain direction within a folded protein. Whereas loops are generally flexible sections of chain, turn structures tend to be more rigid and are stabilized by hydrogen bonding or specific side-chain interactions. These turn structures can be highly important, particularly in antiparallel β-sheet structures, where a complete reversal of the chain is required to enable packing of adjacent strands. Other less frequently occurring elements of secondary structure (e.g., 3_{10} or π helices) can also be identified on the Ramachandran plot.

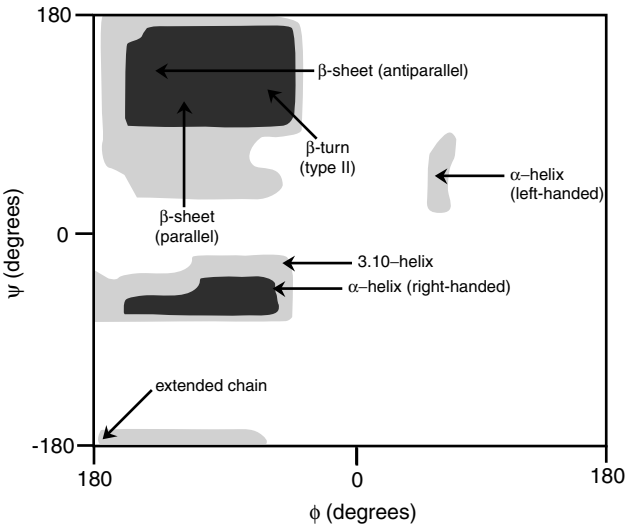


Figure 1.3. A schematic representation of the Ramachandran plot.

So far, we have largely ignored the contributions of the amino acid side chains to protein conformation. One obvious consequence of the existence of a variety of different side chains is the dependence of the Ramachandran plots for each particular (φ_i,ψ_i) pair on the identity of the ith amino acid residue. For example, a significantly larger conformational volume is available to glycine as compared to amino acid residues with bulky side chains. Furthermore, different side chains placed at "strategic" locations may exert a significant influence on the stability of the secondary structural elements. We will illustrate this point using the α-helix as an example. All hydrogen bonds in an α-helix are almost parallel to each other (and to the axis of the helix). This highly ordered pattern of hydrogen bonding results in a noticeable dipole moment, with the N-terminal end of the helix being a positive pole. Obviously, the presence of a positively charged residue at or near the N-terminal end of the helix will destabilize it due to the unfavorable charge–permanent dipole interaction (Eq. 1-2-2). On the other hand, the presence of a negatively charged residue will be energetically favorable and will increase the stability of the helix. Likewise, the presence of charged residues at or near the C-terminal end of the helix will also have a significant influence on the stability of this element of secondary structure. Note, however, that uncharged side chains may also be very important determinants of the higher order structure of proteins and polypeptides due to the so-called *hydrophobic interactions*. These will be considered in Section 1.2.4.

1.2.4. Solvent–Solute Interactions, Hydrophobic Effect, Side Chain Packing, and Tertiary Structure

The term *hydrophobic effect* (16–19) refers to a tendency of nonpolar compounds (e.g., nonpolar amino acid side chains, Table 1.1) to be sequestered from polar solutions (e.g., aqueous solution) into an organic phase. Such behavior is ubiquitous in nature and has been observed and described at least 2 millennia ago, although the term hydrophobic was coined only in 1915 (18). The initial view of the hydrophobic interaction was rather simplistic and implied attraction between like media (e.g., oil–oil attraction). A very different view, which is now commonly accepted, was proposed in the mid-1930s by Hartley, who suggested that nonpolar species are excluded from polar solvent because of their inability to compete with the strong interaction between the polar molecules themselves (20). In

Tanford's words, "*antipathy between hydrocarbon and water rests on the strong attraction of water for itself*" (21). An intriguing aspect of the hydrophobic interaction is that the placement of a hydrocarbon molecule in water may be enthalpically favorable. This fact was the basis for a widespread skepticism over the concept of hydrophobic interactions, although such views did not prevail (22). It is now understood that solvent–solute affinity is determined by the free energy (not the enthalpy alone), and it is the unfavorable free energy that leads to the observed disaffinity of water and nonpolar solutes.

Various microscopic explanations of the hydrophobic effect are usually based on the *frozen water patches* or *microscopic iceberg* model proposed originally by Frank and Evans (23). They suggested that placing a nonpolar solute in water creates a loose "cage" of first-shell water molecules around it. The creation of such a cage has a significant entropic price due to the forced ordering of water, hence the overall unfavorable free energy (despite a favorable enthalpic term). Readers interested in a more detailed account of the physics of hydrophobicity and related phenomena are referred to an excellent tutorial by Southall, Dill, and Haymet (18).

Although the initial work on the hydrophobic effect was focused on hydrocarbons, its main results and conclusions can be easily extended to nonpolar side chains of polypeptides and proteins, which are buried into a *hydrophobic core* of a folded or collapsed protein molecule in order to eliminate, or at least minimize, any contacts with the polar solvent. A very interesting historical account of the elucidation of the nature of the hydrophobic interaction and its role in protein folding can be found in an excellent review by Tanford (24). Hydrophobic side chains are generally more stable if sequestered away from the solvent in protein cores. Proteins tend to be very well-packed molecules so the side-chain atoms sequestered from the solvent must come into close contact with each other, hence the term *hydrophobic packing*. At the same time, hydrophilic residues usually decorate the solvent-exposed surface of the protein. This decoration is achieved by combining the elements of secondary structure (α-helices, β-sheets, and turns) in a unique 3D arrangement, or *tertiary structure*. It is the tertiary structure that affords proteins their unique biological function, whether it be purely structural, the precise spatial organization of side chains to effect catalysis of a reaction, presentation of a surface or loop for signaling or inhibition, creating a cavity or groove to bind ligand, or any of the other vast range of functions that proteins can perform.

Hydrophobic interaction is, of course, not the only driving force giving rise to a unique tertiary structure. Additional stabilization is afforded by the close proximity of acidic and basic residues, which is frequently observed in the folded structure, enabling the formation of salt bridges. These can be viewed as charge–charge interactions (Eq. 1-2-1). We have already mentioned that certain elements of secondary structure have intrinsic (permanent) dipole moments. Favorable arrangement of such dipoles with respect to one another (e.g., in the so-called helical bundles) may also become a stabilizing factor (Eq. 1-2-2) in addition to the hydrophobic interaction. It is probably worth mentioning that in the vast majority of proteins, the interactions stabilizing the tertiary structure are cooperative. In other words, significant enthalpic gains are achieved only if several segments of the protein are in close proximity and interact with each other. All such factors have been evolutionarily optimized for each protein, but the important thing to realize is that any one natural protein sequence has only a single most stable conformation, and the genetically encoded primary sequence alone is necessary and sufficient to define the final folded structure of the protein (Fig. 1.4) (25).

Many proteins adopt similar common structural motifs resulting from combinations of secondary structure elements, such as the alternating $\beta\alpha\beta$ structure, 4-helix bundles, or β-barrels. As more and more protein structures are solved, the number of protein architectures increases, although it has been predicted that there are a limited number of fold motifs (26–30). This conclusion is based on the observations that (1) topological arrangements of the elements of secondary structure are highly skewed by favoring very few common connectivities and (2) folds can accommodate unrelated sequences [as a general rule, structure is more robust than sequence (31,32)]. Therefore, the fold universe appears to be dominated by a relatively small number of giant attractors, each accommodating a large number of unrelated sequences. In fact, the total number of folds is estimated to be <2000, of which 500 have been already characterized. Figure 1.5 represents the 15 most populated folds selected on the basis of a structural annotation of proteins from completely sequenced genomes of 20 bacteria, 5 Archaea, and 3 eukaryotes (33).

The existence of a "finite set of natural forms" in the protein world has inspired some to invoke the notion of Platonic forms that are "determined by natural law" (34), a suggestion that seems more poetic than explanatory. What has become clear though is that very similar tertiary structures can be adopted by quite dissimilar primary sequences (33). Protein primary sequences can be aligned and regions identified that are identical or homologous (meaning the chemical nature of the amino acid side chain is similar, e.g., polar, nonpolar, acidic, basic). However, even sequences with quite low homology can have a very similar overall fold, depending on the tertiary interactions that stabilize them. Although tertiary structure is sometimes viewed as the highest level of spatial organization of single-chain (i.e., monomeric) proteins, an even higher level of organization is often seen in larger proteins (generally,

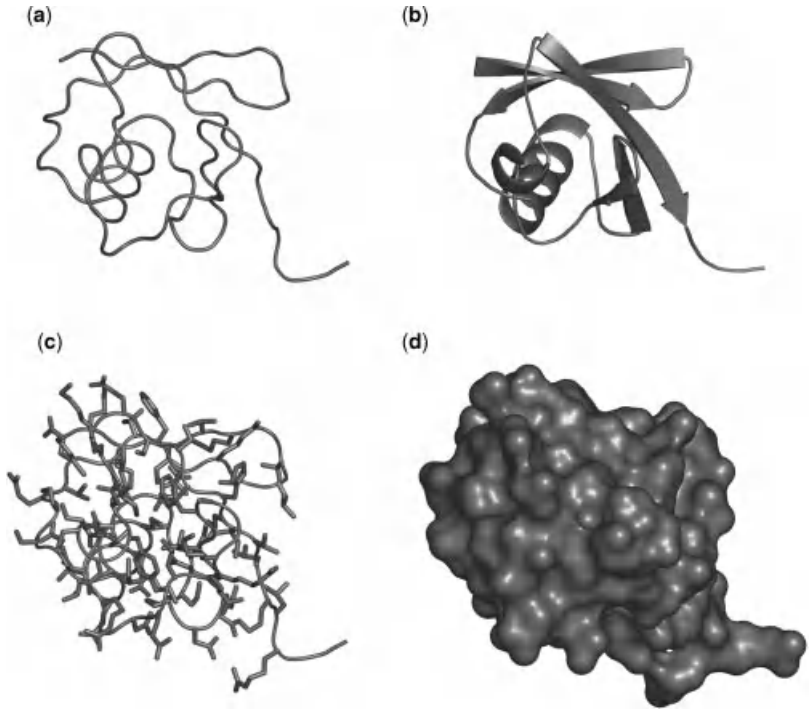


Figure 1.4. Different representations of the higher order structure of natively folded proteins.

>150 amino acid residues). Such proteins form clearly recognizable *domains*, which tend to be contiguous in primary structure and often enjoy a certain autonomy from one another.

1.2.5. Intermolecular Interactions and Association: Quaternary Structure

Above and beyond the folding of monomeric chains, many protein chains can also assemble to form multisubunit complexes, ranging from relatively simple homodimers (example are hemoglobin molecules of primitive verterbrates, e.g., lamprey and hagfish) to large homooligomers (e.g., the iron storage protein ferritin, comprised of 24 identical subunits) to assemblies of different proteins (e.g., ribosomes). Such assemblies are usually considered to be the highest level of molecular organization at the microscopic level, which is usually referred to as *quaternary structure*. Although covalent links are sometimes formed between the monomeric constituents of a multimeric protein assembly (e.g., in the form of disulfide bonds), the noncovalent interactions (discussed in the preceding sections) are usually much more important players.

The archetype of quaternary structure is mammalian hemoglobin, which is a noncovalent tetramer ($\alpha_2\beta_2$) consisting of two pairs of similar monomeric chains (α- and β-globins). The arrangement of monomers in the tetramer, which is in fact a dimer composed of two heterodimers, is crucial for the function of hemoglobin as an oxygen transporter. A tetramer composed of four identical globins (β_4) can also be formed and is indeed present in the blood of people suffering from some forms of thalassemia. However, this homotetramer (termed hemoglobin H or HbH), lacks the most important characteristic of the "normal" hemoglobin (HbA), namely, high cooperativity of oxygen binding.

1.3. THE PROTEIN FOLDING PROBLEM

1.3.1. What is Protein Folding?

Polymers can adopt different conformations in solution depending on functionality and the interaction with neighboring chains, other parts of the same chain, and the bulk solvent. However, almost all synthetic copolymers (i.e., polymers consisting of more than one type of repeat unit) consist of a range of different length chains and, in many cases, a nonspecific arrangement of monomer groups. On the other hand, the primary structure of a given protein is always the same, creating a homogeneous and highly monodisperse copolymer. Protein sequences are generally optimized to prevent nonspecific intermolecular interactions and individual molecules will fold to adopt a unique stable conformation governed solely by the primary sequence of amino acids. The ability of proteins to attain a unique higher order structure sets them apart from most random copolymers. Most proteins can fold reversibly *in vitro*, without being aided by any sophisticated cellular machinery (e.g., chaperones, which we will consider in

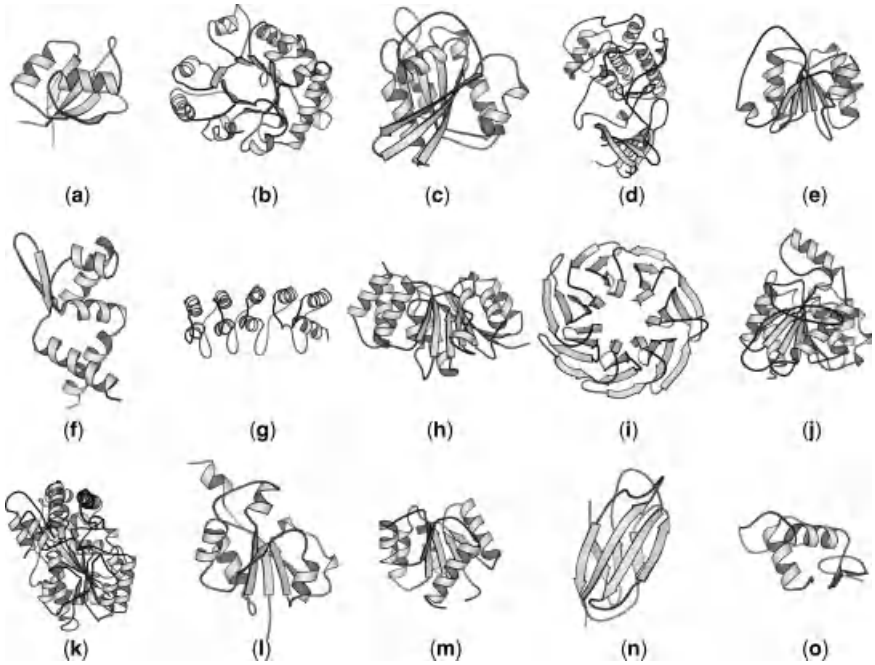


Figure 1.5. The 15 most populated folds selected on the basis of a structural annotation of proteins from the completely sequenced genomes of 20 bacteria, 5 Archaea, and 3 eukaryotes. From left to right and top to bottom, they are ferredoxin-like (4.45%) (*a*), TIM-barrel (3.94%) (*b*), P-loop containing nucleotide triphosphate hydrolase (3.71%) (*c*), protein kinases (PK) catalytic domain (3.14%) (*d*), NAD(P) (nicotinamide adenine dinucleotide phosphate)-binding Rossmann-fold domains (2.80%) (*e*), (deoxyribonucleic acid: ribonucleic acid) (DNA:RNA) binding 3-helical bundle (2.60%) (*f*), α–α superhelix (1.95%) (*g*), *S*-adenosyl-L-methionine-dependent methyltransferase (1.92%) (*h*), 7-bladed β-propeller (1.85%) (*i*), α/β-hydrolases (1.84%) (*j*), PLP-dependent transferase (1.61%) (*k*), adenine nucleotide α-hydrolase (1.59%) (*l*), flavodoxin-like (1.49%) (*m*), immunoglobulin-like β-sandwich (1.38%) (*n*), and glucocorticoid receptor-like (0.97%) (*o*). The values in parentheses are the percentages of annotated proteins adopting the respective folds. [Reprinted from (33). Copyright © 2001 with kind permission Springer Science+Business Media.]

Chapter 9), suggesting that the folding mechanism is solely determined by the primary structure of the protein, as well as the nature of the solvent. Folded proteins may remain stable indefinitely in most cases, suggesting that the native structures represent the global free energy minima among all kinetically accessible states (35).

Two classic puzzles are usually considered in connection with protein folding: (1) the *Blind Watchmaker's paradox* and (2) the *Levinthal paradox*. The former is named after a classic book by Dawkins (36), an outspoken critic of the *intelligent design* concept (37). It states that biological (function-competent) proteins could not have originated from random sequences. The Levinthal paradox states that the folded state of a protein cannot be found by a random search (38). Both paradoxes have been historically framed in terms of a random search through vast spaces (sequence space in the *Blind Watchmaker's paradox* and conformational space in the *Levinthal's paradox*), and the vastness of the searched space is equated with physical impossibility. Both paradoxes are elegantly solved within the framework of the *energy landscape* description of the folding process by invoking the notion of a guided search (39). The concept of protein energy landscapes and its relevance to the protein folding problem will be considered in some detail in Section 1.4.

1.3.2. Why Is Protein Folding So Important?

First, one question is Why do we need to understand protein folding? In the post-genomic era, structure determination has become of paramount importance since it leads to a 3D picture of each gene product, and in many cases gives hints as to the function of the protein. However, the static structure only represents the end point of the chemical reaction of protein folding. Polypeptide chains are translated as extended structures from RNA on the ribosome of cells, but How does this unstructured sequence fold into its final biologically active structure? Are specific local structures present in the newly translated chain? Is there a specific pathway or reaction coordinate of protein folding?

The principles that govern the transitions of biopolymers from totally unstructured to highly ordered states, which often include several subunits assembled in a highly organized fashion, remain one of the greatest mysteries in structural biology (40,41). Deciphering this code is key to understanding a variety of biological processes at the molecular level (recognition, transport, signaling and biosynthesis, etc.), since the specificity of biological activity in proteins, as well as other biomolecules, is dictated by their higher order structure.

Aside from the obvious academic interest to biophysicists in discovering exactly how these biological machines work, there are many more practical implications. Only if we understand all of the processes that are involved in producing a biologically active protein can we hope to harness this power by designing proteins with specific functions. It may already be possible computationally to model an ideal binding site or even optimal arrangement of side chains to catalyze a chemical reaction, but without a thorough knowledge of how this site can be placed into an intact protein molecule, we cannot take advantage of the cellular machinery for the design of therapeutic protein drugs, or even molecules that can catalyze otherwise difficult chemical reactions. For instance, there are many enzymes in nature that catalyze reactions with extremely high specificity and efficiency, whereas chemists lag far behind. Hydrogenase enzymes, for example, catalyze the reduction of protons to produce diatomic hydrogen, a reaction that in a laboratory environment requires application of harsh reactants at elevated temperature or pressure, but that within the catalytic center of the protein occurs at physiological temperatures and with remarkably small energy requirements. Obviously, biological organisms have had a much longer time to optimize these processes relative to the chemical industry. If one can understand in detail the roles of each residue in a protein chain for both the folding and dynamics of the molecule, then the possibilities for protein engineering are boundless. Interestingly, manmade sequences quite often lead to proteins that either do not fold at all or are only marginally stable. This result clearly demonstrates the extremely fine balance of forces present, which can be destroyed by just a single amino acid residue substitution, deletion, or insertion.

Another important aspect of understanding protein folding is to find ways of preventing the process from going awry (42–45). An ever-increasing number of pathological conditions that result from misfolding of proteins in the cell are being identified (46–50). Amyloid plaques actually result from the undesirable formation of quaternary structure when a normally monomeric peptide folds incorrectly and self-assembles to form long proteinaceous fibers. Similarly, other proteins, that are not correctly folded may not present the correct binding surface for interaction with their physiological partners. Thus not only correct folding, but also the correct assembly of proteins, is key to their correct biological function. Even relatively few mutations within a protein sequence may prevent folding to the native structure, and hence prove pathological. In other cases, mutation can reduce the efficiency of folding, or favor an alternative mode of folding that leads to aggregation and deposition of insoluble amyloid plaques within cells. We will consider the issues related to misfolding and aggregation later.

Finally, one more fundamental problem related to protein folding that has become a focal point of extensive research efforts is the prediction of the native structure and function of a protein based on its primary structure. Since the sequence of each natural protein effectively encodes a single tertiary structure, prediction of the latter is, in essence, a global optimization problem, which is similar to one encountered in crystallography and the physics of clusters (51). The complication that arises when such a global optimization methodology is applied to determine the position of the global energy minimum for a protein is the vastness of the system that precludes calculations based on first principles. So far, the most successful methods of structure prediction rely on the identification of a *template protein* of known structure, whose sequence is highly homologous to that of the protein in question. If no template structure can be identified, *de novo* prediction methods can be used, although it remains to be seen if such methods can predict structures to a resolution useful for biochemical applications (52). Prediction of protein function based on its sequence and structure is an even more challenging task, since homologous proteins often have different functions (53).

1.3.3. What Is the Natively Folded Protein and How Do We Define a Protein Conformation?

Before proceeding further with a description of protein folding it would be useful to define some terms commonly used in the field in order to avoid confusion. First, the *native state* of a protein is defined as the fully folded biologically active form of the molecule. This has generally been considered as a single state with a well-defined tertiary structure, as determined by crystallography or nuclear magnetic resonance (NMR) spectroscopy. More recently, researchers have come to appreciate the importance of dynamics within the protein structure. Even the native state is not a static single structure, but may in fact, depending on the protein, have small or even large degrees of flexibility that are important for its physiological function.

Unfortunately, the use of the term *protein conformation* in the literature has become rather inconsistent and often results in confusion. Historically, protein conformation referred to a specific "three-dimensional arrangement of its constituent atoms" (54). This definition, however, is rather

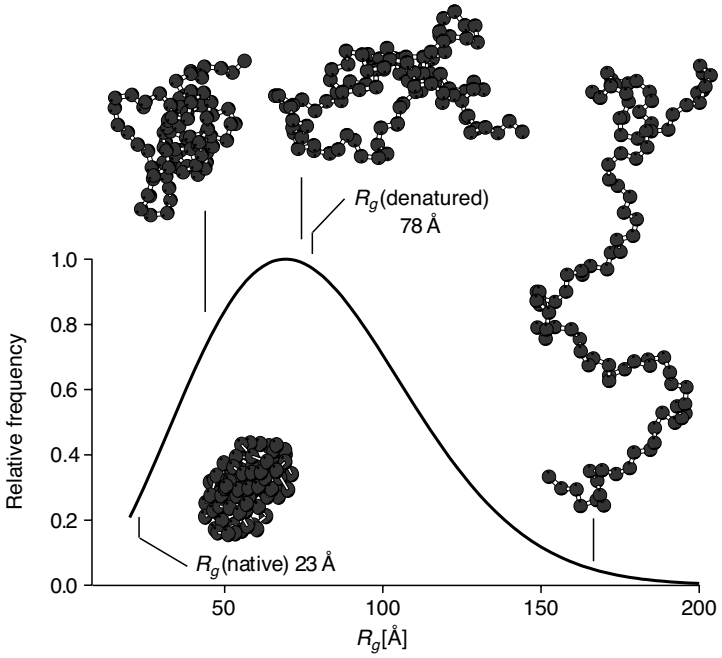


Figure 1.6. Representative configurations of a random coil (a freely joined chain of 100 hard spheres) and the distribution of its radius of gyration R_g. The R_g values of a model protein phosphoglycerate kinase are indicated for comparison. [Adapted from (55) Copyright © 1996 with kind permission from Elsevier.]

narrow, since it does not reflect adequately the dynamic nature of proteins. One particularly annoying complication that arises when conformation is defined using only microscopic terms (e.g., atomic coordinates) is due to the fact that a majority of proteins have segments lacking any stable structure even under native conditions. These could be either the terminal segments that are often invisible in the X-ray structures or flexible loops whose conformational freedom is often required for a variety of functions ranging from recognition to catalysis. In general, it is more than likely that any two randomly selected natively folded protein molecules will *not* have identical sets of atomic coordinates and, as a result, will not be assigned to one conformation if the geometry-based definition is strictly applied. Therefore, it seems that the thermodynamics-based definition of a protein conformation is a better choice. Throughout this book, we will refer to the protein conformation not as a specific microstate, but as a *macrostate*, which can be envisioned as a collection of microstates separated from each other by low energy barriers ($\leq k_B T$). In other words, if one microstate is accessible from another at room temperature, we will consider them as belonging to one conformation, even if there is a substantial difference in their configurations. According to this view, a protein conformation is a continuous subset of the conformational space (i.e., a continuum of well-defined configurations) that is accessible to a protein confined to a certain local minimum. The utility of this definition becomes obvious when we consider non-native protein conformations, although unfortunately it is not without its own problems.*

1.3.4. What Are Non-Native Protein Conformations? Random Coils, Molten Globules, and Folding Intermediates

In the case of unfolded proteins, which are assumed to be completely nonrigid polypeptide chains, the *random coil* (55), we must consider the ensemble of molecules displaying an impressive variety of configurations (Fig. 1.6). In a truly random coil, as might be the case for a synthetic polymer with identical monomer units in a good solvent, there may well be no conformational preferences for the chain. However, proteins are decorated with side chains of a different chemical nature along their length, such that in water or even in a chemical denaturant one might expect there to be local preferences due to hydrophilic or hydrophobic interactions, and indeed steric effects. Thus for a number of proteins studied in solution, some persistent local and nonlocal conformational effects have been detected,

*For example, this definition is temperature dependent. Indeed, if any two local minima are separated by a high energy barrier ($>k_B T$), the interconversion between these two states does not occur readily at room temperature (T), and these two states should be viewed as two different conformations. However, raising the temperature significantly above room temperature will eventually make passage over this barrier possible, leading to a merging of the two microstates to a single conformation.

indicating that an unfolded protein generally is not in fact a truly random coil. On the other hand, the enthalpy of these interactions is very small in comparison to the entropy of the flexible chain so the overall free energy of each of these conformers will be very similar. On a free energy surface, these would be represented as shallow wells in the generally flat surface of unfolded state free energy.

The relative position of a local energy minimum with respect to the native state gives rise to a further set of descriptions of intermediate states. As a protein folds it may sample stabilizing conformations that contain persistent structure, constituting a local free energy minimum. At the earliest stages of folding there may be only a few interactions that may be very transient: These are termed *early intermediates*. By contrast, species may accumulate further along in the folding process that contain a large although incomplete number of native-like contacts. These are referred to as *late intermediates*, implying that they should form toward the end of the kinetic folding process. There is also the possibility that these local minima arise from stabilizing contacts that are not present in the native protein, and in fact need to be disrupted before the molecule can productively fold. These *off-pathway intermediates* may also arise from intermolecular interactions between folding chains and can lead to nonproductive aggregation that prevents further folding.

The above intermediate states form during folding in the "forward" direction from the unfolded to the native state and, since they are only partially stable, generally do not accumulate sufficiently to be detected other than transiently. It is also possible that such intermediates may form during the reverse process, that is protein unfolding, allowing them to be studied by other methods. Unfortunately, the conditions for unfolding (e.g., chemical denaturant, low pH, high temperature) are generally so harsh that once the stabilizing interactions in the native state have been removed, the unfolding process occurs with high cooperativity and without accumulation of intermediates. However, under mildly denaturing conditions, partially folded states have been detected at equilibrium for a number of proteins, and these have been termed *molten globules* (56). The original definition of the molten globule state was quite rigid: a structural state that has significant secondary structure, but with no fixed tertiary interactions. There are various biophysical tests for this, such as the ability of the protein to bind hydrophobic dyes, consistent with a significant amount of exposed hydrophobic surface area, as would be expected for a partially folded state. The definition has become somewhat relaxed to include many other partially folded ensembles observed, kinetically or at equilibrium, which almost fit the definition. What is clear is that the molten globule itself is a much more dynamic structure than previously thought. Several new concepts have been introduced to reflect the structural diversity and dynamic character of the molten globule state, such as "a precursor of the molten globule" and "a highly structured molten globule" (57).

One common question that arises is whether the equilibrium molten globule intermediate is actually the same species as that detected in the folding pathway of proteins. Thermodynamically there is nothing to suggest they should be, since the equilibrium by definition is independent of the pathway (58,59). However, comparisons of the characteristics of transient intermediates with the corresponding equilibrium partially folded state have concluded that the similarities are very close, at least for the proteins studied (60–63). Also, a number of states transiently populated by the native state ensemble under mildly destabilizing conditions have been shown to have similarities to folding intermediates. Thus it seems likely that, at least in the later stages of folding, there is indeed some kind of folding–unfolding pathway with specific intermediate states visited in both the folding and unfolding directions.

1.3.5. Protein Folding Pathways

In Section 1.3.4, we began to use the term *folding pathway*, which is understood to be a series of structural changes leading from the fully denatured state of the protein to its native conformation. Introduction of the concept of a folding pathway resolves the Levinthal paradox mentioned earlier by suggesting that the folding process is a directed process involving conformational biases, rather than a merely random conformational search. Despite the vast number of degrees of freedom in macromolecules, the number of folding pathways was initially believed to be rather limited (64). A general scheme of protein folding within this paradigm is presented in terms of rapid equilibration of unfolded protein molecules between different conformations prior to complete refolding. Such equilibria favor certain compact conformations that have lower free energies than other unfolded conformations, and some of these favored conformations are important for efficient folding. The rate-limiting step is thought to occur late in the pathway and to involve a high-energy, distorted form of the native conformation. The latter is a single transition state through which essentially all molecules refold (65).

The classic folding pathway paradigm specifically states that "*proteins are not assembled via a large number of independent pathways, nor is folding initiated by a nucleation event in the unfolded protein followed by rapid growth of the folded structure*" (65). Nevertheless, a large body of experimental evidence now suggests the existence of a large number of folding routes. Furthermore, over the past several years it has become clear that the length of the polypeptide chain is an important factor in determining mechanistic details. Smaller proteins (< 100 residues) appear to prefer a nucleation-type mechanism that does not involve any specific metastable intermediate species (66). On the other

hand, for a number of larger proteins, intermediate states with specific regions of early formed stable structure have been established. If an intermediate is detected, then this argues strongly for a specific protein folding pathway, but in the protein lysozyme, for instance, parallel folding pathways were found, suggesting multiple possible trajectories. In smaller proteins, no observable intermediates accumulate, which might indicate a less directed folding mechanism, or indeed that partially formed structures are too labile to be detected. Based on these observations, a modified view of a folding pathway invokes several common stages of folding, consistent with the progressive development of structure and stability through an ever-slowing set of reactions (67).

Three major models have arisen to attempt to explain this narrowing of the conformational search process (68). The *simple framework* model suggests that secondary structure elements would first form based on their sequence-intrinsic propensities, followed by collision of these preformed structures to form tertiary interactions. Alternatively, *nucleation* of short regions of sequence to form transient secondary structure could act as a template upon which adjacent parts of the chain would condense and propagate structure. The third, *hydrophobic collapse* mechanism calls for the hydrophobic residues to conglomerate nonspecifically to minimize solvent exposure, followed by rearrangement to the final native structure. The actual mechanism of protein folding probably involves some or all of these processes and there is evidence for each mechanism from folding studies of different proteins.

The controversy in regard to whether protein folding follows a specific pathway or whether each molecule follows a completely different trajectory to achieve its final folded state has been elegantly resolved in the so-called *new view of protein folding* (69–71). The centerpiece of this theory is the concept of a protein energy landscape of conformational space, which is discussed in Section 1.4.

1.4. PROTEIN ENERGY LANDSCAPES AND THE FOLDING PROBLEM

1.4.1. Protein Conformational Ensembles and Energy Landscapes: Enthalpic and Entropic Considerations

Classically, a simple chemical reaction is considered to proceed along a reaction coordinate, in which chemical bonds are formed or broken in a well-defined manner, with a transition state at the highest point on the energy profile where bond breaking and formation is occurring, and possibly detectable metastable intermediate structures at local free energy minima. Transition state theory can be applied to relate reaction rates to the heights of the various free energy barriers along the reaction coordinate. Macromolecules are much more complex, however, and the folding of a protein involves the formation of a large number of interactions that may be either local in nature or indeed involve regions that are quite distant in the polypeptide sequence. Nevertheless, for many years the protein folding reaction was assumed to occur via a similar sequential pathway, perhaps involving a number of intermediate species along the way, but for each unfolded molecule the mechanism of folding to the native state was identical.

Significant advances in theory during the past decade have changed our understanding of the basic principles that govern the protein folding process and have offered an elegant way to resolve the Levinthal paradox (70,72–76). In the case of small organic molecules in a reaction, there are only a small number of conformations available to the reactant species, but for an unfolded protein the conformational space sampled by the unstructured chain is vast by comparison, even for a relatively small protein. Thus it might seem difficult to imagine the chain becoming oriented in such a way as to proceed to fold via a single pathway, and the process would surely be extremely inefficient. Realization that folding may proceed through multiple parallel pathways, rather than a single route, has led to introduction of the concepts of protein energy landscapes (or folding funnels), a cornerstone of the "new view" of protein folding.

Protein folding can be viewed much like any other chemical reaction, which may be represented in 3D by a conformational energy surface: The trajectories on these surfaces lead from reactants (unstructured states) to products (the native state). Because entropy plays a much more significant role in protein folding reactions, it is necessary to consider the free energy, rather than simply potential energy (77). The enthalpic gain of forming hydrogen bonds and making favorable hydrophobic or hydrophilic contacts is compensated by a significant loss of entropy as the chain becomes more and more conformationally restricted. For simple molecules, the entropic term is generally far less significant, but in the case of a folding protein the overall free energy of stabilization in the folded protein may be only a few kilocalories per mole (kcal/mol), being the very small difference between large ΔH (formation of stabilizing interactions) and $T\Delta S$ (loss of entropy upon folding to the native state) terms. The conformational entropy loss for a protein, which continues to adopt a more well-defined 3D structure, is often defined on a "per residue" basis. It can be estimated by using only the backbone entropy [the entropy loss due to side chain packing is significantly less (75)]:

$$\Delta s = s_{\mathrm{u}} - s_{\mathrm{f}} \approx k_{\mathrm{B}} \cdot \ln\left(\frac{\Omega_{\mathrm{u}}}{\Omega_{\mathrm{f}}}\right) \qquad (1\text{-}4\text{-}1)$$

where $s_{\mathrm{u,f}}$ and $\Omega_{\mathrm{u,f}}$ represent the entropies and the numbers of microstates per residue in the unfolded and natively folded forms of the protein, respectively. Estimations of Δs for small proteins at room temperature give an entropy loss

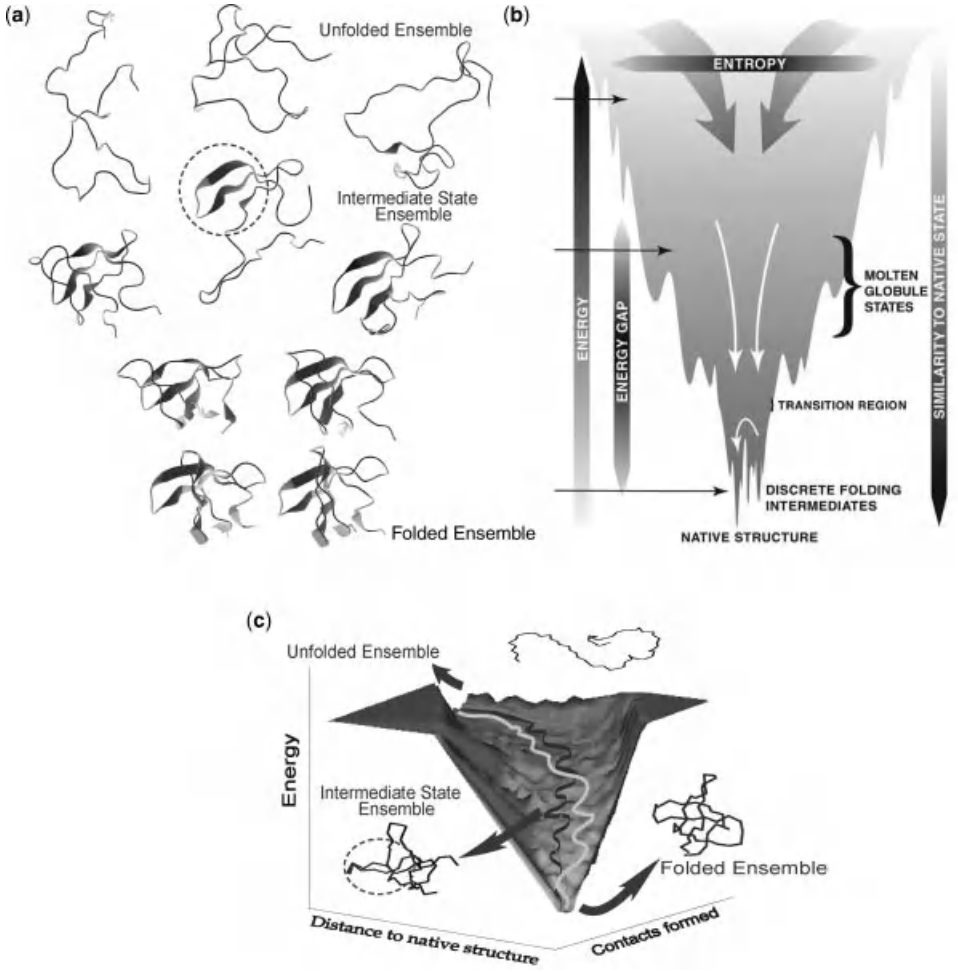


Figure 1.7. Schematic representation of a protein folding funnel. As the large ensemble of structures of an unfolded polypeptide compacts, forming native-like contacts through intermediate states and finally to the native state (*a*), the energy surface can be schematically represented as seen in panel (*b*). Multiplicity of folding routes is shown with different folded trajectories on the energy surface (*c*). However, asymmetry of the surface biases the trajectories toward the preferred route, which can be considered a folding pathway. [Reprinted with permission from (78). Copyright © 2001 C. Clementi and G. Bamberg.]

on the order of tens of joules per mole kelvin residue [J/(mol·K·residue)] (75). To ensure fast folding despite the unfavorable entropy change, the corresponding free energy surface (or energy landscape) must have the form of a multidimensional "funnel" (Fig. 1.7), where the vertical axis (the depth of the funnel) represents the number of native contacts made (Q) or the relative free energy of the conformational space. The horizontal axis corresponds to the conformational entropy of the system. In this representation, it becomes clear that as the folding chain makes more native contacts, the chain entropy is reduced along with the overall free energy. The native state resides at the bottom of the potential well, characterized by low entropy and a global free energy minimum. A funneled energy landscape is robust to both environmental changes and sequence mutations, since most potentially competing low-energy states are similar in structure (75), as represented by the multiple local minima at the bottom of the funnel in Figure 1.7.

An important feature of the folding funnel is that its slopes are not always monotonic, hence the competition between the "downhill slide" toward the native fold and the possibility of equally favorable excursions into local free energy minima, depending on the ruggedness of the energy surface. These local minima may represent transient formation of partially folded species, accumulation of intermediates or indeed misfolded forms, depending on the trajectory taken by the chain along the energy surface. This general model allows for multiple pathways with no specific order of structure formation, but subtle changes in the energy surface would lead to a far more directed approach by energetically favoring particular regions of conformational space. It also provides at least one possible solution

for Levinthal's problem of folding time scale. If a favored conformational space were visited on the energy landscape, then it would be sufficiently stabilized to reside there for a longer time, and hence restrict the conformational search for the most stable native structure.

As already stated, one of the most important features that distinguishes these folding funnel models from more traditional reaction coordinates is the key feature of conformational entropy. Thus we are forced to consider each stage along a folding reaction not as a single structure, but instead as a conformational ensemble. The unfolded state (the nearly flat plain at the top of the funnel) demonstrates the large number of microstates in the unfolded polypeptide chain, whereas the deep energy well of the native state may be a single structure or indeed a small ensemble of similar structures closely related in energy. Along the folding trajectory as the chain becomes more ordered, this ensemble become smaller and smaller, but throughout there is always some conformational flexibility that must be considered.

It must be stressed that these folding landscapes are purely theoretical. Based on experimental data it has proved possible to style different landscapes more as a representation than a true physical picture, showing local free energy wells for intermediate species, and alternate possible routes for parallel pathways. Even with the most powerful modern computing facilities, however, it is not yet possible to predict the folding pathway(s) of a protein, although some trajectories in a computational ensemble may appear to fit well with experimental data (79). Only with a complete understanding of the energetics involved in driving a protein to fold will we be able to computationally predict how a given protein will fold.

Semiquantitative models and experiments are revealing how the folding free energy surface is sculpted by the protein sequence and its environment. Although any downhill path from the unfolded state will eventually lead to a native conformation at the bottom of the funnel, the asymmetry (energetic heterogeneity) of the surface can bias the choice of folding routes (80). The existence of such "preferred" folding routes can be observed experimentally and interpreted in terms of folding pathways (see above). The sometimes conflicting demands of folding, structure, and function determine which folding pathways, if any, dominate (76).

1.4.2. Equilibrium and Kinetic Intermediates on the Energy Landscape

Under native conditions, folding of small proteins usually appears to be a highly cooperative process (81,82). By using standard biophysical techniques, only the native and unfolded states are generally sufficiently populated to be detectable. For instance, a titration of chemical denaturant studied by circular dichroism (CD) for most proteins will generate a series of spectra that can be deconvoluted to contributions solely from the two end points of the unfolding reaction, namely, the native and denatured states. This finding led to the belief that the folding of proteins was not only cooperative, but also two-state, that is, without populated intermediate states.

In contrast, under certain conditions, significant accumulation of an intermediate conformational ensemble may occur if the free energy barrier is sufficiently high, referred to as an *equilibrium intermediate*. These species can be studied in great detail since the rate of conversion is low, allowing significant structural information to be obtained (83). Classic examples of these include the partially folded state of the apo-form of myoglobin (84), acid- and alcohol-induced *A*-state of ubiquitin (85), or the acid-induced molten globule of α-lactalbumin (86). As already mentioned, these equilibrium intermediate states in some cases may represent important conformations visited along the folding trajectory. Therefore, structural information about these states can give valuable clues as to the nature of the conformational search process. Transient formation or indeed accumulation of certain intermediates is usually induced *in vitro* by simply changing the protein's environment (by varying the solution pH, temperature, presence of chaotropes, etc.) (61,87). This change can result in significant alterations of the energy surface, decreasing the free energy of the intermediate states, and thus increasing their equilibrium population. One needs to be aware, however, that such equilibrium intermediates may differ significantly from the kinetically observed species. As pointed out by Fersht and co-workers (59), an equilibrium intermediate need not by definition be on the preferred kinetic folding pathway since thermodynamic (equilibrium) parameters are independent of mechanism. The goal here though is to make these elusive states more amenable to study using a variety of biophysical techniques, in order to determine not only the conformational preferences of a partially folded protein, but also the possible conformations transiently visited by the native state that may be vital to its *in vivo* function.

Refolding of large proteins often does not conform to the simple two-state model discussed in the beginning of this section. Nevertheless, such proteins may be spontaneously refolded by rapid dilution from a chemically denatured state into native conditions. Kinetic studies of these processes have enabled detection of transient *kinetic intermediates*, which serve as "resting points" in the protein folding process. Since the energy difference between the global minimum and any of the surrounding local minima is usually quite high, the Boltzmann weight of the states that correspond to the local minima is very low. Under native conditions, kinetic intermediates become populated only transiently during refolding experiments. These species have been the focus of close experimental scrutiny, since their structure and behavior may reveal many intimate details of the protein folding process. In order to be detected,

these species must have characteristics different from either the native or unfolded states. For instance, the aromatic residues may experience an environment that makes them hyperfluorescent, or a secondary structure may form in advance of tertiary interactions making the intermediate detectable using stopped-flow optical techniques (e.g., CD). Alternatively, the secondary structure may protect certain amide protons against exchange with bulk solvent, which can be detected by NMR or mass spectrometry (MS), as we will see in the following chapters. The experimental identification and characterization of kinetic intermediates has been the focus of a great deal of research over the past couple of decades, as researchers attempt to glean the determinants of protein folding. Due to their transient nature, however, kinetic intermediates often cannot be observed directly and their properties can only be inferred from indirect measurements.

If the energy barrier separating a kinetic intermediate from the native conformation is high, a significant accumulation of such intermediates may occur. These are referred to as *kinetically trapped* or *metastable* intermediate states. Under certain conditions, excessive accumulation of metastable kinetic intermediates in the course of the folding process may trigger nonspecific interactions among them and, in extreme cases, aggregation. Likewise, aggregation can also be caused by incorrect folding (e.g., due to a sequence mutation). Aggregation processes *in vivo* are prevented by chaperones, a special class of proteins that bind and sequester the misfolded and partially folded polypeptides (88–91). It is important to note, however, that the chaperones only assist, but do not in themselves direct protein folding.

1.5. PROTEIN DYNAMICS AND FUNCTION

1.5.1. Limitations of the Structure–Function Paradigm

Proteins carry out their functions by interacting with other proteins, as well as other molecules ranging from giant biopolymers (e.g., DNA) to small organic molecules and monatomic ligands. In all cases, protein–ligand binding is the first stage of the interaction, which can be followed by a variety of processes ranging from sophisticated chemical transformation of the ligand (e.g., enzyme catalysis) to simple release of the ligand in the presence of other cofactors (e.g., transport proteins). In this section, we will only consider the main characteristics of the protein–ligand binding process.

Binding has been traditionally considered within a framework of the structure–function paradigm, a cornerstone of molecular biology for many years. It was >100 years ago that Fischer coined the term *lock-and-key* to emphasize the requirement for a stereochemical fit between an enzyme and its substrate in order for binding to occur (92). The limitations of this view of the binding process became obvious in the middle of the twentieth century, when a large body of newly acquired information on enzyme kinetics appeared to be in conflict with the notion that "*the enzyme was a rather rigid negative of the substrate and that the substrate had to fit into this negative to react*" (93). The revision of the lock-and-key theory by Koshland (92) led to a rise of the so-called *induced fit* theory, whose major premise was that the "*reaction between the enzyme and substrate can occur only after a change in protein structure induced by the substrate itself*". Conformational changes occurring during enzyme catalysis are relatively small scale and affect mostly the catalytic site (94). Similarly, the conformational changes occurring in other proteins as a result of induced fit-type binding usually affect a limited fraction of the protein structure. An example of this is ferric ion binding by the iron-transport protein transferrin, an event that results in the repositioning of two protein domains within each lobe of the protein (Fig. 1.8). Although the overall effect of such repositioning is quite significant (and results in closing the cleft between the two domains), the number of affected amino acid residues does not exceed a dozen (95). More recently, numerous examples of large-scale conformational changes induced by ligand binding have been reported. The most extreme case is represented by the so-called *intrinsically disordered* proteins, which actually lack stable structure under native conditions in the absence of the ligand (96).

The above considerations strongly suggest that structure is not the sole determinant of protein function. As elegantly put by Onuchic and Wolynes (80), "*the twentieth century's fixation on structure catapulted folding to center stage in molecular biology. The lessons learned about folding may, in the future, increase our understanding of many functional motions and large-scale assembly processes*". In Section 1.5.2, we will consider various aspects of protein dynamics under native conditions that may be important modulators or even determinants of function.

1.5.2. Protein Dynamics Under Native Conditions

With very few exceptions, protein structure under native conditions is not a rigid crystalline state, but undergoes local breathing motions, involving anything from side-chain rotation to rearrangement of secondary structure elements relative to each other. Although the existence of such motions within the native state of the protein can be detected with a variety of experimental techniques, their exact nature remains the subject of discussion in the literature. The commonly accepted models of local dynamics within natively folded proteins invoke the notions of *structural fluctuation* (localized transient unfolding affecting only few atoms within the protein) (97) or a *mobile defect,* which

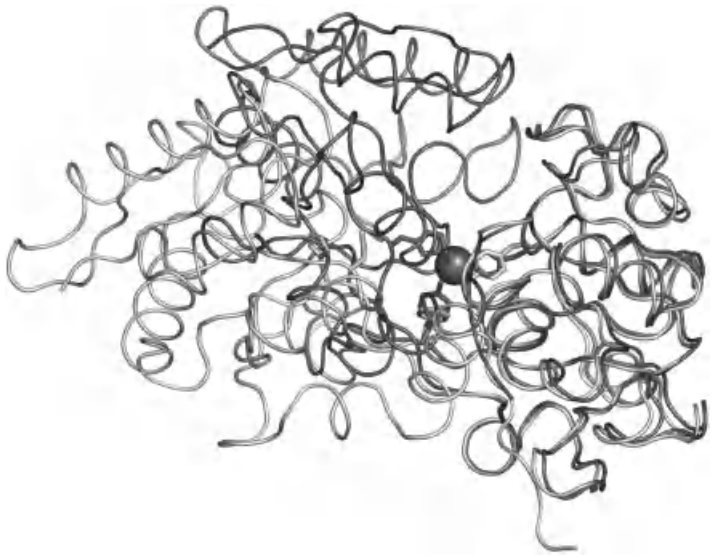


Figure 1.8. Superimposed crystal structures of the apo and holo forms of the *N*-lobe of human serum transferrin.

considers not only the emergence and dissipation of local disorder, but also the possibility of its propagation through the protein structure. Although the latter model has not enjoyed as much attention as the former, thorough theoretical considerations suggest that local perturbations of the secondary structure may in fact propagate through certain elements of the secondary structure (e.g., α-helices) in the form of a soliton (98). Alternatively, local dynamics can be described with a *solvent penetration* model (slow diffusion of the solvent molecules into and out of the protein interior) (99). Such description is actually very similar to the mobile defect model, as applied to the integral solute–solvent system, instead of the protein molecule alone.

Above and beyond local structural fluctuations, the dynamics of proteins under native conditions is exemplified by transiently sampling alternative (higher energy or "activated") conformations. Such activated (non-native) states are often functionally important despite their low Boltzmann weight (100,101). An example of such behavior can be seen in cellular retinoic acid binding protein I (CRABP I), which sequesters and transports insoluble all-*trans* retinoic acid (RA) in the cytosol. The structures of the apo and the holo forms of this protein are very similar, consistent with the lock-and-key type of binding. However, the native structure of CRABP I provides no clue as to how the ligand gets access to the internal protein cavity, which is its binding site (Fig. 1.9). Obviously, in order to provide entrance into the cavity, a fraction of the native structure has to be lost transiently, an event consistent with the notion of sampling an activated protein state. Realization of the importance of transient non-native protein structures for their function has not only greatly advanced our understanding of processes as diverse as recognition, signaling, and transport, but also has had profound practical implications, particularly for the design of drugs targeting specific proteins (102).

Many proteins use dynamics as a means of communication between different domains. This process, by which a signal, such as a binding event in one domain triggers a conformational change in another domain, is known as *allostery*. The paradigm for this effect is hemoglobin, a tetrameric protein mentioned earlier in this chapter. Binding a molecule of oxygen at the heme site of the α-chain induces a change in the oxygen affinity of the β-chain binding site by rearrangement of interdomain interactions (103,104). Another example is the chaperone protein DnaK, which assists in preventing the misfolding of nascent chains as they emerge from the ribosome. This 70-kDa protein consists of an ATPase domain joined via a short linker region to a peptide-binding domain. Binding of adenosine triphosphate (ATP) causes a conformational change in the peptide-binding domain that increases its affinity for substrate. Subsequent hydrolysis of nucleotide in the ATPase domain signals a conformational change in the adjacent domain that releases the unfolded polypeptide and allows it to begin to refold. The exact mechanism by which this allosteric communication occurs is still poorly understood. It is clear, however, that it must involve dynamic events at the interdomain interface that transmits the signal between the two binding sites (105).

1.5.3. Is Well-Defined Structure Required for Functional Competence?

A very interesting class of proteins that came to prominence in the early 2000s relies on dynamics even more heavily

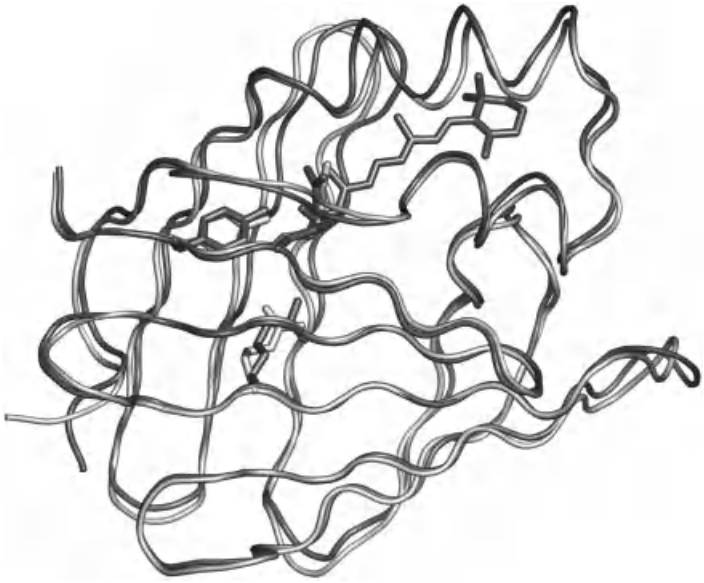


Figure 1.9. Overlaid crystal structures of the apo- and holo- forms of cellular retinoic acid binding protein I.

compared to the examples considered in Section 1.5.2. *Intrinsically disordered* proteins are remarkable in that they appear to have very little stable folded structure in isolation, contrary to our classical view of proteins as folded species. Several hundred proteins have now been identified that contain large segments of disorder even under native conditions (106), and many of these proteins seem to serve a wide variety of functions *in vivo* (107–109). A number of these are involved in activation or inhibition of transcription or translation, and while these proteins appear unstructured under native conditions they undergo a structural transition in the presence of their cognate substrate, whether this be another protein or a recognition site on a molecule of DNA or RNA (96,110). One important aspect of intrinsic disorder may be the necessity for this class of proteins to recognize and bind to multiple sites. Whereas a highly structured protein may only have a limited and very specific binding site available to it, one that has a highly dynamic structure should be able to adapt to a variety of different structural motifs. This intrinsic disorder phenomenon also seems contrary to the paradigm that an unstructured protein should be targeted for proteolysis, or else degradation by the proteasome. It seems this class of proteins manages to avoid such scenarios either by having regions that are sterically inaccessible or else do not contain residues that are sensitive to proteases. Indeed one other observation is that many intrinsically disordered proteins have a relatively short lifetime in the cell: They are expressed as needed in response to a signal and then rapidly removed by degradation. This would provide an efficient mechanism to switch on or off a cellular process for only a short period of time. A larger number of proteins in the eukaryotic genome have been predicted to have disordered regions compared to prokaryotes, perhaps indicating the need for higher organisms to adjust more rapidly to environmental changes.

1.5.4. Biomolecular Dynamics and Binding from the Energy Landscape Perspective

The development of the folding funnel concept also has far-reaching consequences for our understanding of how proteins interact with each other and with other ligands. One theory offered as a general scheme of protein folding and binding implies that the only difference between the two processes is chain connectivity (111,112), namely, that monomeric protein folding represents an energy funnel for a single chain, whereas protein–protein association and peptide binding is a similar landscape, but with discontinuous backbone connections. In the more general case, however, the concept can be extended to encompass the chemical nature of the ligand and the energetics of the binding process, whether it be a noncovalent interaction or a chemical process as in the case of enzymatic catalysis; the energy funnel concept can be applied theoretically to describe the process by which a protein recognizes and binds to another molecule. Rigid proteins that bind ligands via a lock-and-key type mechanism presumably do not require significant dynamic events, so they will have few local minima similar in energy to the native state. In contrast, those proteins that utilize an induced fit binding mechanism may have a rugged energy surface characterized by a number of local minimum conformational states at the bottom of the folding–binding funnel (111).

The idea of a binding funnel has also been demonstrated computationally by Zhang et al. (113) to explain the fast

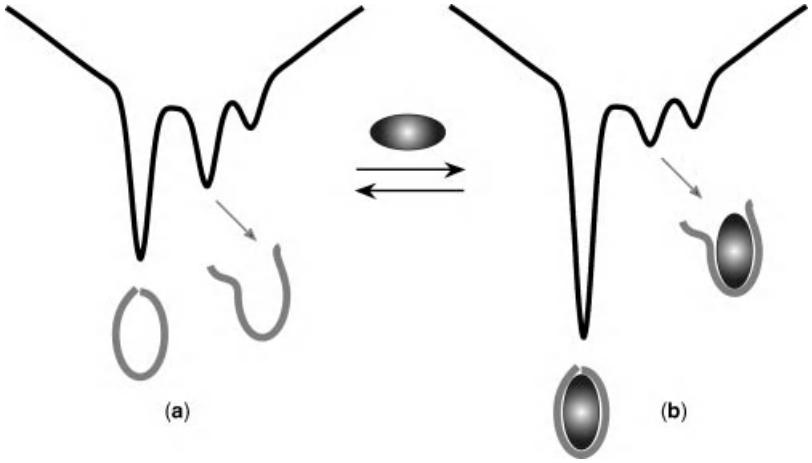


Figure 1.10. Schematic representations of the energy landscapes for the apo (*a*) and holo (*b*) forms of a protein whose ligand-binding behavior conforms to the lock-and-key type interaction.

protein–ligand association rates exhibited by many proteins. In this model, the initial collision event is accompanied by favorable interactions to form a long-lived encounter complex that significantly limits the search process to the ligand-bound conformation. The funnel energy landscape of protein binding may be a common feature in protein–protein associations (114,115). Knowledge of the relative energies and structural features of the local conformational minima available to proteins is clearly key to the understanding of what makes proteins efficient in binding their physiological ligands. Likewise, an extension of the protein folding problem is to understand how protein monomers assemble as functional multimers or other macromolecular assemblies. In the cellular environment, this must be an efficient process aided by dynamics and specific recognition events, all of which may be described in terms of the energetics of accessible conformational space.

All visibly different modes of binding (lock-and-key, induced fit, and binding of intrinsically disordered proteins) appear to have only quantitative differences within the framework provided by the binding funnel concept. This point is illustrated in Figures 1.10–1.12, which represent hypothetical folding funnels for proteins of each class in the absence and presence of their respective ligands. A protein whose ligand-binding behavior conforms to the lock-and-key type interaction (e.g., CRABP I considered earlier) is suggested to have an activated state whose structural features increase the rate of ligand entry into the binding site (Fig. 1.10). The protein molecules sample this activated state relatively frequently due to its relatively low energy. Once the ligand enters the binding site, its interaction with the protein increases the stability of the native conformation. Although "visitations" to the activated state are still possible, they do not occur as frequently due to the increased energy difference between the two states. As a result, the protein can acquire the ligand relatively easy, but does not release it unless the energy landscape is altered again, (e.g., by a competing receptor of the ligand).

A similar analysis can be carried out for proteins conforming to the induced-fit type behavior (we will use the N-lobe of human serum transferrin as an example). Although the X-ray data suggest the existence of two distinct conformations of the protein depending on the presence of the ligand (open conformation for the apo form and closed for the holo form of the protein), there is experimental evidence suggesting that both conformations coexist in solution in equilibrium (116). The open conformation is, of course, favored in the absence of the ligand, while iron binding shifts the equilibrium toward the closed state (Fig. 1.11). Such a shift is qualitatively similar to the one considered for the lock-and-key interaction. The only difference is that the protein state corresponding to the global energy minimum in the absence of the ligand becomes "downgraded" to the status of an activated state (local energy minimum) as a result of the ligand binding.

Finally, folding of an intrinsically unstructured protein in the process of ligand binding can be viewed as a preferential stabilization of one particular conformation among many available to the protein in the ligand-free form (Fig. 1.12). An example of such behavior is presented by the β-chain of mammalian hemoglobins, which populate at least four different states (only one of them appears to be very close to the compact natively folded conformation) in solution in the absence of its binding partner, α-globin (117). Again, the general features of binding in this scheme appear to be very similar to those seen in the previous two examples (Figs. 1.10 and 1.11), the major distinct feature being the absence of the preferred conformation in the absence of the ligand.

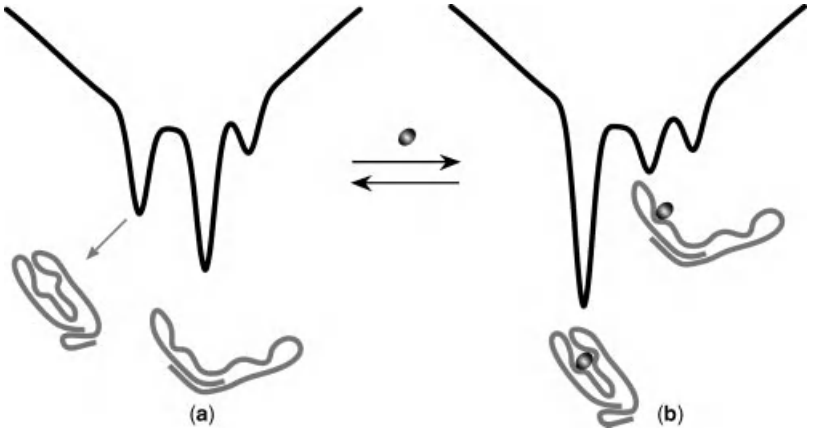


Figure 1.11. Schematic representations of the energy landscapes for the apo (*a*) and holo (*b*) forms of a protein whose ligand-binding behavior conforms to the induced fit type interaction.

1.5.5. Energy Landscapes Within a Broader Context of Nonlinear Dynamics: Information Flow and Fitness Landscapes

It becomes increasingly clear that the significance of the concept of energy landscapes extends well beyond the fields of protein folding and even molecular biophysics. Huang and Ingber (118) questioned the validity of the commonly accepted paradigm of cell regulation as a collection of pathways that link receptors with genes by asking "*Can identification of all these signaling proteins and their assignment into distinct functional pathways lead to the full understanding of developmental control and cell fate regulation?*" The existence of distributed information within cellular signaling, as well as the fact that a single signaling molecule may activate several genes and even produce opposite effects depending on its microenvironment, led to a suggestion that the concept of *linear signaling pathways* is inappropriate. The suggestion that the signaling and regulatory pathways are not simple linear connections between receptors and genes is similar to the earlier realization of the limits of the classic concept of protein folding pathways. The paradox of signaling nonlinearity is resolved by the introduction of a concept of *cellular states*, and the switches between these states are viewed as *biological phase transitions* (118). In this new view, cell fates are viewed as common end programs or attractors within the entire regulatory network, which can be visualized as a *potential landscape* with multiple minima. Each minimum corresponds to a certain fate of the cell (e.g., differentiation, proliferation, apoptosis). A similar idea was recently proposed as the basis of a quantitative model of carcinogenesis, in which normal cells *in vivo* occupy a ridge-shaped maximum in a well-defined *tissue fitness landscape*, a configuration that allows cooperative coexistence of multiple cellular populations (119).

Finally, *dynamic fitness landscapes* have proven to be a valuable concept in evolutionary biology, molecular evolution (120,121), combinatorial optimization, and the physics

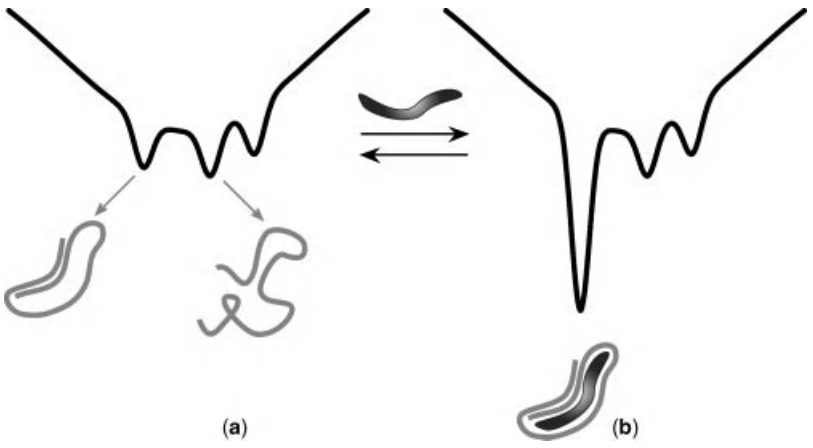


Figure 1.12. Schematic representations of the energy landscapes for the apo (*a*) and holo (*b*) forms of an intrinsically unstructured protein whose folding is induced by the ligand.

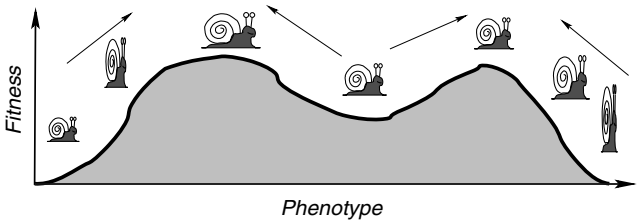


Figure 1.13. Schematic representation of a fitness landscape, where the fitness of a snail species depends on its shape. Mutations continuously produce variants that are selected if their fitness is larger than the fitness of the current "wild-type" snail. As a consequence, the shape of the snails changes over time until it reaches a maximum of the fitness landscape. [Reprinted with permission from (122). Copyright © 2002 Society for Industrial and Applied Mathematics. All rights reserved.]

of disordered systems (122). In evolutionary biology, this concept is often used to visualize the relationship between genotypes (or phenotypes) and replication success, an idea initially put forward by Wright (123). An evolving population typically climbs uphill in the fitness landscape, until it reaches a local optimum (Fig. 1.13), where it then remains, unless a rare mutation opens a path to a new fitness peak.

1.6. PROTEIN HIGHER ORDER STRUCTURE AND DYNAMICS FROM A BIOTECHNOLOGY PERSPECTIVE

In the first edition of this book, the discussion of protein folding and conformation was focused primarily on fundamental aspects, and the goal was to understand some general principles of protein (and, more broadly, biopolymer) behavior *in vitro* with the hope of being able to also apply this knowledge to *in vivo* situations. Spectacular progress has been made in the field of biotechnology in the past several years, which has resulted in a dramatic explosion of both the number of protein-based drugs and the range of the diseases they can treat (124,125), bringing to the forefront another very important aspect of protein folding and dynamics. Indeed, the large size of protein therapeutics (from several kDa to nearly 1 MDa, well beyond the molecular weight range of classical small molecule medicinal drugs) leads to an important distinction between small molecule drugs (where covalent structure is the sole determinant of the 3D structure and, ultimately, the therapeutic properties of the drug) and protein pharmaceuticals (where the large physical size makes a multitude of noncovalent contacts not only inevitable but, in fact, the defining element of their 3D structure). Correct folding is vital not only for the ability of a protein to execute its biological function, but also for many other aspects of its behavior (126). Failure to fold or maintain the native conformation obviously has a negative impact on the efficacy of the protein drug, since the recognition of a range of physiological targets requires that the native conformation be maintained throughout the lifecycle of a protein molecule. Proteins that are not folded properly are prone to aggregation both *in vitro* and *in vivo*, and are targeted by proteases both inside and outside the cell, which obviously impacts bioavailability of the protein drug. Furthermore, misfolding and aggregation may trigger an immune response, thereby adversely affecting the safety profile of the protein drug. Critical dependence of the protein drug's potency, stability, and safety on conformation makes its characterization an essential element throughout the drug development process from design to manufacturing to postapproval monitoring. Throughout this book, we will provide several examples of how MS can be used to probe various aspects of conformation, dynamics, and stability of protein pharmaceuticals, as well as their interactions with physiological partners and therapeutic targets.

REFERENCES

1. Thieffry, D. and Sarkar, S. (1998) Forty years under the central dogma, *Trends Biochem. Sci. 23*, 312–316.
2. Nirenberg, M. (2004) Historical review: Deciphering the genetic code—a personal account, *Trends Biochem. Sci. 29*, 46–54.
3. Hatfield, D. L. and Gladyshev, V. N. (2002) How selenium has altered our understanding of the genetic code, *Mol. Cell. Biol. 22*, 3565–3576.
4. Ibba, M. and Soll, D. (2002) Genetic code: Introducing pyrrolysine, *Curr. Biol. 12*, R464–R466.
5. Budisa, N., Minks, C., Alefelder, S., Wenger, W., Dong, F., Moroder, L., and Huber, R. (1999) Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire, *FASEB J. 13*, 41–51.
6. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science 301*, 964–967.
7. Avetisov, V., and Goldanskii, V. (1996) Mirror symmetry breaking at the molecular level, *Proc. Natl. Acad. Sci. USA 93*, 11435–11442.
8. Podlech, J. (2001) Origin of organic molecules and biomolecular homochirality, *Cell. Mol. Life Sci. 58*, 44–60.
9. Warshel, A. and Papazyan, A. (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling, *Curr. Opin. Struct. Biol. 8*, 211–217.
10. Schutz, C. N. and Warshel, A. (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models?, *Proteins 44*, 400–417.
11. Harvey, S. C. (1989) Treatment of electrostatic effects in macromolecular modeling, *Proteins 5*, 78–92.
12. Daune, M. (1999) *Molecular biophysics: structures in motion*, Oxford University Press, Oxford, New York.
13. Mirsky, A. E. and Pauling, L. (1936) On the structure of native, denatured, and coagulated proteins, *Proc. Natl. Acad. Sci. USA 22*, 439–447.

14. Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins—2 hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. USA 37*, 205–211.
15. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations, *J. Mol. Biol. 7*, 95–99.
16. Tanford, C. (1978) Hydrophobic effect and organization of living matter, *Science 200*, 1012–1018.
17. Scheraga, H. A. (1998) Theory of hydrophobic interactions, *J. Biomol. Struct. Dyn. 16*, 447–460.
18. Southall, N. T., Dill, K. A., and Haymet, A. D. J. (2002) A view of the hydrophobic effect, *J. Phys. Chem. B 106*, 521–533.
19. Widom, B., Bhimalapuram, P., and Koga, K. (2003) The hydrophobic effect, *Phys. Chem. Chem. Phys. 5*, 3085–3093.
20. Hartley, G. S. (1936) *Aqueous solutions of paraffin-chain salts; a study in micelle formation*, Hermann & Cie, Paris.
21. Tanford, C. (1979) Interfacial free-energy and the hydrophobic effect, *Proc. Natl. Acad. Sci. USA 76*, 4175–4176.
22. Hildebrandt, J. H., Nemethy, G., Scheraga, H. A., and Kauzmann, W. (1968) A criticism of term hydrophobic bond, *J. Phys. Chem. 72*, 1841–1842.
23. Frank, H. S. and Evans, M. W. (1945) Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures; partial molar entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes, *J. Chem. Phys. 13*, 507–532.
24. Tanford, C. (1997) How protein chemists learned about the hydrophobic factor, *Protein Sci. 6*, 1358–1366.
25. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223–230.
26. Chothia, C., Hubbard, T., Brenner, S., Barns, H., and Murzin, A. (1997) Protein folds in all-α and all-β classes, *Annu. Rev. Biophys. Biomol. Struct. 26*, 597–627.
27. Zhang, C. and DeLisi, C. (1998) Estimating the number of protein folds, *J. Mol. Biol. 284*, 1301–1305.
28. Govindarajan, S., Recabarren, R., and Goldstein, R. A. (1999) Estimating the total number of protein folds, *Proteins 35*, 408–414.
29. Hou, J., Sims, G. E., Zhang, C., and Kim, S.-H. (2003) A global representation of the protein fold space, *Proc. Natl. Acad. Sci. USA 100*, 2386–2390.
30. Caetano-Anolles, G. and Caetano-Anolles, D. (2003) An evolutionarily structured universe of protein architecture, *Genome Res. 13*, 1563–1571.
31. Rost, B. (1997) Protein structures sustain evolutionary drift, *Fold. Des. 2*, S19–S24.
32. Wood, T. C. and Pearson, W. R. (1999) Evolution of protein sequences and structures, *J. Mol. Biol. 291*, 977–995.
33. Zhang, C. and DeLisi, C. (2001) Protein folds: molecular systematics in three dimensions, *Cell. Mol. Life Sci. 58*, 72–79.
34. Denton, M. J., Marshall, C. J., and Legge, M. (2002) The protein folds as platonic forms: New support for the pre-Darwinian conception of evolution by natural law, *J. Theor. Biol. 219*, 325–342.
35. Epstein, C. J., Goldberger, R. F., and Anfinsen, C. B. (1963) Genetic control of tertiary protein stucture—studies with model systems, *Cold Spring Harbor Symp. Quant. Biol. 28*, 439–439.
36. Dawkins, R. (1996) *The blind watchmaker: why the evidence of evolution reveals a universe without design*, Norton, New York.
37. Pennock, R. T. (2001) *Intelligent design creationism and its critics: philosophical, theological, and scientific perspectives*, MIT Press Cambridge, MA.
38. Levinthal, C. (1968) Are there pathways for protein folding, *J. Chim. Phys. Phys. Chim. Biol. 65*, 44–45.
39. Dill, K. (1999) Polymer principles and protein folding, *Protein Sci. 8*, 1166–1180.
40. Radford, S. E. and Dobson, C. M. (1999) From computer simulations to human disease: emerging themes in protein folding., *Cell 97*, 291–298.
41. Jaenicke, R. (1998) Protein self-organization in vitro and in vivo: partitioning between physical biochemistry and cell biology., *Biol. Chem. 379*, 237–243.
42. Jaenicke, R. (1995) Folding and association versus misfolding and aggregation of proteins., *Philos. Trans. R. Soc. London Ser. B 348*, 97–105.
43. Rochet, J.-C. and Lansbury, P. T. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol. 10*, 60–68.
44. Uversky, V. N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go?, *Cell. Mol. Life Sci. 60*, 1852–1871.
45. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation, *Semin. Cell Dev. Biol. 15*, 3–16.
46. Carrell, R. W. and Lomas, D. A. (1997) Conformational disease, *Lancet 350*, 134–138.
47. Bellotti, V., Mangione, P., and Stoppini, M. (1999) Biological activity and pathological implications of misfolded proteins, *Cell. Mol. Life Sci. 55*, 977–991.
48. Thompson, A. J. and Barrow, C. J. (2002) Protein conformational misfolding and amyloid formation: characteristics of a new class of disorders that include Alzheimer's and Prion diseases, *Curr. Med. Chem. 9*, 1751–1762.
49. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature* (*London*), *416*, 507–511.
50. Caughey, B. and Lansbury, P. T. (2003) Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders, *Annu. Rev. Neurosci. 26*, 267–298.
51. Wales, D. J. and Scheraga, H. A. (1999) Global optimization of clusters, crystals, and biomolecules, *Science 285*, 1368–1372.
52. Schonbrun, J., Wedemeyer, W. J., and Baker, D. (2002) Protein structure prediction in 2002, *Curr. Opin. Struct. Biol. 12*, 348–354.

53. Whisstock, J. C. and Lesk, A. M. (2003) Prediction of protein function from protein sequence and structure, *Q. Rev. Biophys. 36*, 307–340.
54. Price, N. C. (2000) Conformational issues in the characterization of proteins, *Biotechnol. Appl. Biochem. 31*, 29–40.
55. Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996) The concept of a random coil. Residual structure in peptides and denatured proteins, *Fold. Des. 1*, R95–R106.
56. Ptitsyn, O. B. (1995) Molten globule and protein folding, *Adv. Protein Chem. 47*, 83–229.
57. Uverskii, V. N. (1998) [How many molten globules states exist?], *Biofizika 43*, 416–421.
58. Clarke, J. and Fersht, A. R. (1996) An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway, *Fold. Des. 1*, 243–254.
59. Clarke, J., Itzhaki, L. S., and Fersht, A. R. (1997) Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways?, *Trends Biochem. Sci. 22*, 284–287.
60. Jennings, P. A. and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, *Science 262*, 892–896.
61. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native state hydrogen exchange., *Science 269*, 192–196.
62. Bai, Y. (1999) Equilibrium amide hydrogen exchange and protein folding kinetics, *J. Biomol. NMR 15*, 65–70.
63. Parker, M. J. and Marqusee, S. (2001) A kinetic folding intermediate probed by native state hydrogen exchange, *J. Mol. Biol. 305*, 593–602.
64. Creighton, T. E. (1984) Pathways and mechanisms of protein folding, *Adv. Biophys. 18*, 1–20.
65. Creighton, T. E. (1988) Toward a better understanding of protein folding pathways, *Proc. Natl. Acad. Sci. USA 85*, 5082–5086.
66. Gunasekaran, K., Eyles, S. J., Hagler, A. T., and Gierasch, L. M. (2001) Keeping it in the family: folding studies of related proteins, *Curr. Opin. Struct. Biol. 11*, 83–93.
67. Matthews, C. R. (1993) Pathways of protein folding, *Annu. Rev. Biochem. 62*, 653–683.
68. Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding?, *Trends Biochem. Sci. 28*, 18–25.
69. Baldwin, R. L. (1995) The nature of protein folding pathways: the classical versus the new view., *J. Biomolec. NMR 5*, 103–109.
70. Dill, K. A. and Chan, H. S. (1997) From Levinthal to pathways to funnels., *Nat. Struct. Biol. 4*, 10–19.
71. Pande, V. S., Grosberg, A., Tanaka, T., and Rokhsar, D. S. (1998) Pathways for protein folding: is a new view needed?, *Curr. Opin. Struct. Biol. 8*, 68–79.
72. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways and the energy landscape of protein folding: a synthesis., *Proteins 21*, 167–195.
73. Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: the energy landscape perspective, *Annu. Rev. Phys. Chem. 48*, 545–600.
74. Brooks, C. L., 3rd, Gruebele, M., Onuchic, J. N., and Wolynes, *P. _G.* (1998) Chemical physics of protein folding, *Proc. Natl. Acad. Sci. USA 95*, 11037–11038.
75. Plotkin, S. S. and Onuchic, J. N. (2002) Understanding protein folding with energy landscape theory. Part I: Basic concepts, *Q. Rev. Biophys. 35*, 111–167.
76. Gruebele, M. (2002) Protein folding: the free energy surface, *Curr. Opin. Struct. Biol. 12*, 161–168.
77. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment., *Trends Biochem. Sci. 25*, 331–339.
78. Brooks, C. L., 3rd, Onuchic, J. N., and Wales, D. J. (2001) Statistical thermodynamics. Taking a walk on a landscape, *Science 293*, 612–613.
79. Fersht, A. R. and Daggett, V. (2002) Protein folding and unfolding at atomic resolution, *Cell 108*, 573–582.
80. Onuchic, J. N. and Wolynes, P. G. (2004) Theory of protein folding, *Curr. Opin. Struct. Biol. 14*, 70–75.
81. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Fold. Des. 3*, R81–91.
82. Myers, J. K. and Oas, T. G. (2002) Mechanisms of fast protein folding, *Annu. Rev. Biochem. 71*, 783–815.
83. Ptitsyn, O. B. (1994) Kinetic and equilibrium intermediates in protein folding., *Protein Eng. 7*, 593–596.
84. Eliezer, D., Jennings, P. A., Dyson, H. J., and Wright, P. E. (1997) Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR, *FEBS Lett. 417*, 92–96.
85. Brutscher, B., Bruschweiler, R., and Ernst, R. R. (1997) Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ^{1}H, ^{13}C, and ^{15}N nuclear magnetic resonance spectroscopy, *Biochemistry 36*, 13043–13053.
86. Baum, J., Dobson, C. M., Evans, P. A., and Hanley, C. (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin, *Biochemistry 28*, 7–13.
87. Bai, Y., Englander, J. J., Mayne, L., Milne, J. S., and Englander, S. W. (1995) Thermodynamic parameters from hydrogen exchange measurements., *Methods Enzymol. 259*, 344–356.
88. Fink, A. L. (1999) Chaperone-mediated protein folding, *Physiol. Rev. 79*, 425–449.
89. Hartl, F. U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein, *Science 295*, 1852–1858.
90. Trombetta, E. S. and Parodi, A. J. (2003) Quality control and protein folding in the secretory pathway, *Annu. Rev. Cell Dev. Biol. 19*, 649–676.
91. Mogk, A. and Bukau, B. (2004) Molecular chaperones: structure of a protein disaggregase, *Curr. Biol. 14*, R78–R80.
92. Fischer, E. (1894) Einfluss der configuration auf die wirkung derenzyme, *Ber. Dt. Chem. Ges. 27*, 2985–2993.

93. Koshland, D. E. (1958) Application of a theory of enzyme specificity to protein synthesis, *Proc. Natl. Acad. Sci. USA 44*, 98–104.
94. Koshland, D. E., Jr. (1998) Conformational changes: how small is big enough?, *Nat. Med. 4*, 1112–1114.
95. Jeffrey, P. D., Bewley, M. C., MacGillivray, R. T., Mason, A. B., Woodworth, R. C., and Baker, E. N. (1998) Ligand-induced conformational change in transferrins: crystal structure of the open form of the N-terminal half-molecule of human transferrin, *Biochemistry 37*, 13978–13986.
96. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573–6582.
97. Maity, H., Lim, W. K., Rumbley, J. N., and Englander, S. W. (2003) Protein hydrogen exchange mechanism: local fluctuations, *Protein Sci. 12*, 153–160.
98. d'Ovidio, F., Bohr, H. G., and Lindgard, P. A. (2003) Solitons on H bonds in proteins, *J. Phys. Condens. Mat. 15*, S1699–S1707.
99. Miller, D. W. and Dill, K. A. (1995) A statistical mechanical model for hydrogen exchange in globular proteins, *Protein Sci. 4*, 1860–1873.
100. Tsai, C. D., Ma, B., Kumar, S., Wolfson, H., and Nussinov, R. (2001) Protein folding: binding of conformationally fluctuating building blocks via population selection, *Crit. Rev. Biochem. Mol. Biol. 36*, 399–433.
101. Papoian, G. A. and Wolynes, P. G. (2003) The physics and bioinformatics of binding and folding-an energy landscape perspective, *Biopolymers 68*, 333–349.
102. Carlson, H. A. (2002) Protein flexibility and drug design: how to hit a moving target, *Curr. Opin. Chem. Biol. 6*, 447–452.
103. Perutz, M. F. (1972) Stereochemical mechanism of cooperative effects in haemoglobin, *Biochimie 54*, 587–588.
104. Perutz, M. F., Wilkinson, A. J., Paoli, M., and Dodson, G. G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited, *Annu. Rev. Biophys. Biomol. Struct. 27*, 1–34.
105. Pellecchia, M., Montgomery, D. L., Stevens, S. Y., Vander Kooi, C. W., Feng, H. P., Gierasch, L. M., and Zuiderweg, E. R. (2000) Structural insights into substrate binding by the molecular chaperone DnaK, *Nat. Struct. Biol. 7*, 298–303.
106. Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder in the Protein Data Bank, *J. Biomol. Struct. Dyn. 24*, 325–342.
107. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks, *FEBS J. 272*, 5129–5148.
108. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophys. J. 92*, 1439–1456.
109. Dunker, A. K., Oldfield, C., Meng, J., Romero, P., Yang, J., Chen, J., Vacic, V., Obradovic, Z., and Uversky, V. (2008) The unfoldomics decade: an update on intrinsically disordered proteins, *BMC Genomics 9*, S1.
110. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins, *J. Mol. Biol. 323*, 573–584.
111. Tsai, C.-J., Xu, D., and Nussinov, R. (1998) Protein folding *via* binding and *vice versa.*, *Fold. Des. 3*, R71–R80.
112. Tsai, C.-J., Kumar, S., Ma, B., and Nussinov, R. (1999) Folding funnels, binding funnels, and protein function, *Protein Sci. 8*, 1181–1190.
113. Zhang, C., Chen, J., and DeLisi, C. (1999) Protein-protein recognition: exploring the energy funnels near the binding sites, *Proteins 34*, 255–267.
114. Tovchigrechko, A. and Vakser, I. A. (2001) How common is the funnel-like energy landscape in protein–protein interactions?, *Protein Sci. 10*, 1572–1583.
115. Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition, *Nat. Chem. Biol. 5*, 789–796.
116. Baker, H. M., Anderson, B. F., and Baker, E. N. (2003) Dealing with iron: Common structural principles in proteins that transport iron and heme, *Proc. Natl. Acad. Sci. USA 100*, 3579–3583.
117. Griffith, W. P. and Kaltashov, I. A. (2003) Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry, *Biochemistry 42*, 10024–10033.
118. Huang, S. and Ingber, D. E. (2000) Shape-dependent control of cell growth, differentiation and apoptosis: switching between attractors in cell regulatory networks, *Exp. Cell Res. 261*, 91–103.
119. Gatenby, R. A. and Vincent, T. L. (2003) An evolutionary model of carcinogenesis, *Cancer Res. 63*, 6212–6220.
120. Drossel, B. (2001) Biological evolution and statistical physics, *Adv. Phys. 50*, 209–295.
121. Wilke, C. O., Ronnewinkel, C., and Martinetz, T. (2001) Dynamic fitness landscapes in molecular evolution, *Phys. Rep. 349*, 395–446.
122. Reidys, C. M. and Stadler, P. F. (2002) Combinatorial landscapes, *SIAM Rev. 44*, 3–54.
123. Wright, S. (1967) "Surfaces" of selective value, *Proc. Natl. Acad. Sci. USA 58*, 165–172.
124. Walsh, G. (2007) *Pharmaceutical biotechnology: concepts and applications*, John Wiley & Sons, Hoboken, NJ.
125. Walsh, G. (2010) Biopharmaceutical benchmarks 2010, *Nat. Biotech. 28*, 917–924.
126. Korotchkina, L. G., Ramani, K., and Balu-Iyer, S. V. (2008) Folding considerations for therapeutic protein formulations, In *Molecular Biology of Protein Folding*, *Part A*, pp. 255–270, Elsevier Academic Press Inc., San Diego.

2

OVERVIEW OF TRADITIONAL EXPERIMENTAL ARSENAL TO STUDY BIOMOLECULAR STRUCTURE AND DYNAMICS

Biophysicists traditionally use a battery of experimental techniques, which have not until relatively recently included mass spectrometry (MS). The purpose of this chapter is to present an overview of the alternative methodologies that have been routinely employed, explaining the details that can be gleaned from each technique. This will be used as a jump point in future chapters to demonstrate the complementary nature of mass spectrometric methodologies in the biophysical arena.

2.1. X-RAY CRYSTALLOGRAPHY

2.1.1. Fundamentals

Macromolecules are generally too small to be observed directly by optical microscopy because their size is insufficient to diffract light of wavelengths in the visible region of the electromagnetic spectrum. However, the interatomic spacing ($\sim$1.5 Å) is similar to the wavelength of X-rays. In an ordered crystal, the molecules are exactly aligned in one or only few orientations in three-dimensional (3D) space in a repeating array of *unit cells*. The unit cell is defined as the smallest volume element that represents the entire crystal. Adjacent unit cells are related to one another by translation along the cell axes. Thus the entire crystal can be constructed by laying unit cells side by side. Within a protein molecule each atom will have an exactly equivalent atom at the exact same position in a molecule in each and every other unit cell, forming crystal planes of atoms. These planes, or more correctly the planes formed by the electron clouds between atoms, are what produce the diffraction pattern when exposed to X-ray radiation (1).

Lattice planes in crystals will diffract incident X-rays in a coherent manner depending on the angle of the incident beam, according to Bragg's law:

$$2d_{hkl} \sin\theta = n\lambda \qquad (2\text{-}1\text{-}1)$$

which indicates that a beam of X-rays incident at an angle θ on a series of parallel planes separated by spacing d (the indices h, k, and l represent the lattice indices in 3D) will be reflected only if the left side of the equation is an integer multiple of the incident wavelength. Other angles will of course cause reflections, but only if the above relationship holds will the waves reflected by adjacent lattice planes produce constructive interference and emerge from the crystal in phase to produce a strong diffracted beam.

In principle, one could use Eq. 2-1-1 to relate the measured diffraction angles to interatomic distances and from there calculate the molecular structure in the crystal. In practice, however, this is computationally extremely complex, especially for macromolecules containing a large number of atoms. In practice, waves may be considered as periodic functions of the form:

$$f(x) = F\cos 2\pi(hx + \alpha) \qquad (2\text{-}1\text{-}2)$$

In one dimension (1D), the function describing this wave has amplitude F, frequency h, and phase α. Each diffracted X-ray beam may then be considered in terms of a Fourier

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

series of reflections from each atom in the structure, to yield structure factors in 3D, F_{hkl}, of the form:

$$F_{hkl} = f_a + f_b + f_c \cdots \quad (2\text{-}1\text{-}3)$$

The contribution from each atom (*a, b, c*...) is summed to describe the reflection. Expanding into 3D leads to complicated sums of complex numbers to describe the contribution of reflections from every atom in the unit cell to each observed diffraction peak, which leads to a description of the electron density at every point in the unit cell:

$$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}| e^{-2i\pi(hx+ky+lz-\alpha'_{hkl})} \quad (2\text{-}1\text{-}4)$$

This equation represents the electron density map at every point (*x,y,z*) in 3D space in the unit cell, volume *V*. Each F_{hkl} represents a specific reflection, giving rise to a peak in the diffraction pattern. Clearly, solution of this equation is an extremely complex computational problem, made significantly worse by the fact that a complete solution also requires knowledge of the phase (α) of each of the waves, a parameter that cannot be directly obtained from measurement of the diffraction pattern. In fact, the so-called phase problem can be the main bottleneck in crystal structure determination.

For small molecules, crystallization is often a relatively simple matter, but for proteins and other macromolecules growing crystals of sufficient quality and size to obtain high-quality diffraction patterns is a significant challenge. One of the most common approaches to crystallization of proteins is the "hanging drop" method, which involves suspending a droplet of protein solution above a reservoir of buffer in a closed container. At equilibrium the vapor pressure inside the container is equal to that in the droplet, providing conditions under which protein crystals may be formed. In practice, a wide variety of different buffer conditions (varying pH, temperature, protein concentration, and ionic strength, etc.) are assayed to determine optimal conditions. If only small crystals are obtained initially, these may be used as seeds to obtain large high-quality crystals for diffraction. Generally speaking, crystals must be at least 0.5 mm in the shortest dimension to obtain high-quality diffraction data.

Once a crystal is produced it is sealed in a capillary with a small amount of mother liquor (the buffer used to produce the crystal) and mounted on a sample stage (known as a goniometer) that can be precisely rotated in 3D in the sample beam. The sample can then be exposed to the X-ray source and diffraction patterns can be recorded for different orientations of the crystal with respect to the beam. Diffraction patterns at the various angles are recorded using area detectors consisting of arrays of scintillation counters that can count X-ray photons. Crystals are often damaged by heating or free radicals produced from X-ray exposure; consequently a number of individual crystals are generally required for a complete structure determination.

As mentioned above the problem of determining phase is often a significant factor in analyzing the data to obtain a structure. One way around this is to also grow crystals containing heavy atoms (commonly Hg, Pb, or Au), which in many cases can be accommodated within the crystal by binding to the protein at specific sites. Provided that the protein structure and the structure of the crystals are not changed by the presence of these atoms (in this case, the crystal lattice is described as isomorphic), then diffraction patterns from these *heavy atom derivatives* can be used to determine the phase values to a first approximation by a process known as *isomorphous replacement*. Subtracting out the diffraction data for the protein leaves only the reflections produced by the presence of the heavy atoms in the crystal, and since there are relatively few of these atoms it is a significantly easier task to determine their positions in the unit cell. Other approaches, such as anomalous scattering and molecular replacement (using known structures of similar proteins) can also be employed to solve the phase problem, and hence lead to a solution of the crystal structure of the biomolecule.

2.1.2. Crystal Structures at Atomic and Ultrahigh Resolution

Acquisition of crystallographic data at atomic resolution (1.2 Å or higher) provides several important benefits (2). Structure refinement can be done with only weak stereochemical constraints. Mismodeling can be easily corrected or avoided during the refinement, since the atom types can be distinguished at this resolution. Finally, solvent molecules can be placed accurately within the shells (3). At a resolution 0.8 Å or higher, hydrogen atoms become visible and the atoms can no longer be considered spherical particles due to deformation of electron clouds into bonding and nonbonding orbitals. Such visualization of electrons allows the charge distribution and the bond polarization and order to be determined (4), yielding a very informative picture of the protein structure.

2.1.3. Crystal Structures of Membrane Proteins

Membrane proteins constitute a particularly difficult class of biopolymers because of their poor solubility characteristics (a more extensive discussion of properties of membrane proteins will be presented in Chapter 9). Despite some spectacular successes, X-ray crystallography of membrane proteins is still considered an art due to extreme difficulties associated with production of crystals suitable for X-ray diffraction analyses (5,6). An alternative to conventional X-ray crystallographic analysis of a membrane protein is its reconstitution into two-dimensional (2D) membrane protein

crystals in the presence of lipids (7). Unfortunately, most 2D crystals are not sufficiently ordered to provide resolution better than 4 Å (5).

2.1.4. Protein Dynamics and X-Ray Diffraction

Protein structures are represented in crystallography as a set of atoms in 3D space, which are assumed to exhibit harmonic and isotropic vibrations (quantitatively described using B-factors). It has been realized for quite some time that this assumption is an oversimplification that does not provide an adequate description of protein mobility (8). Several methods have been developed to provide more assessment of protein dynamics based on observed crystal heterogeneity [reviewed in (8)]. Despite such progress, information on large-scale dynamic events is usually very difficult to obtain due to the influence of packing forces in protein crystals.

Several crystallographic techniques have been developed in recent years aimed specifically at characterization of protein dynamics by means of X-ray crystallography. It is well known that unlike crystals of small organic molecules, those of proteins contain significant amounts of solvent. Not only do these solvent molecules weaken the intermolecular contacts in the crystal lattice, but they also permit in many cases diffusion of small substrates through the lattice and subsequent binding to the active sites of the enzymes (9,10). These observations have led to the conclusion that crystals of biological macromolecules bear significant resemblance to very concentrated solutions (10). This notion, of course, downplays the fact that large-scale dynamic events are almost always suppressed by crystal-packing forces and simply cannot occur within the framework of the crystal lattice. However, if the macromolecular dynamics does not involve large-scale motions, methods of X-ray crystallography can be used successfully to characterize these dynamic events. In particular, there are numerous examples of using crystallographic methods to study enzyme dynamics by specifically targeting transient intermediate states on the enzymatic pathway. This is usually achieved by either trapping the unstable species (11) or using time-resolved methods (10). The former strategy employs adjustments of temperature, solution pH, and other experimental variables forcing unstable reaction intermediates to accumulate and be stabilized in the protein crystals (11). The latter strategy, on the other hand, exploits both chemical and structural heterogeneity present in the sample throughout the reaction. Such (time-dependent) heterogeneity is then deconvoluted into structures of intermediates using sophisticated methods of data analysis (10).

2.2. SOLUTION SCATTERING TECHNIQUES

Unlike X-ray diffraction, the solution scattering techniques briefly reviewed in this section do not provide structural information on biopolymers at high (atomic) resolution. They do, however, often provide a means to observe large-scale structural heterogeneity of proteins directly in solution. As such, they have become indispensable tools in studies of protein dynamics in solution. In addition, they may complement the diffraction data by providing low-resolution structural information on the protein segments that fail to produce adequate diffraction patterns.

2.2.1. Static and Dynamic Light Scattering

Light scattering has been used historically to characterize synthetic polymers, although in recent years it has been increasingly applied to study biopolymers as well. In *static light scattering*, intensity of scattered light is measured as a function of the scattering vector $\vec{q}$:

$$\mathrm{q} = |\vec{\mathrm{q}}| = \frac{4\pi n}{\lambda} \sin\frac{\theta}{2} \qquad (2\text{-}2\text{-}1)$$

where n is the solvent refractive index, and θ is a scattering angle (12). Intensity of the scattered light adjusted for background scattering and normalized to a reference solvent gives the Rayleigh ratio $\mathrm{R_s(q)}$, which can be expressed in dilute solutions as (12):

$$\frac{1}{\mathrm{R_s(q)}} \cdot \frac{4\pi^2 n^2 \left(\frac{dn}{dc}\right)^2}{N_\mathrm{A}\lambda^4} = \frac{1}{\mathrm{P}(q)\cdot\langle M\rangle} + 2B\mathrm{c} \qquad (2\text{-}2\text{-}2)$$

where c is protein concentration, N_A is the Avogadro's number, $\langle\,\mathrm{M}\,\rangle$ is the weight-averaged molecular weight of protein particles in solution, B is the second virial coefficient describing inter-particle interactions in solution, and $\mathrm{P}(q)$ is a particle shape factor. If particle size is small compared to the light wavelength, then it obviously acts as a "point scatterer", and its shape is irrelevant for scattering (i.e., $\mathrm{P}(q) = 1$ in Eq. 2-2-2). Most proteins fall into this category, hence the only key parameter that can be obtained from static light scattering methods is the molecular weight. Such sensitivity of static light scattering measurements to the average molecular weight of the protein particles makes it very useful in the studies of protein association and aggregation. In the following chapters, we will refer to several examples of such measurement, namely, use of static light scattering to study association–dissociation equilibria of hemoglobin tetramers (13), as well as formation of amyloid fibrils (14). Once the particle size becomes comparable to the wavelength used in the experiments (i.e., radius of gyration $R_g \sim 1/q$), $\mathrm{P}(q)$ can be approximated reasonably well as the quadratic function of qR_g, and the scattering profiles can be used to determine the particle gyration radius (12).

Dynamic light scattering is sensitive to the diffusion of scattering particles in solution, as it measures the intensity

of light scattered at a fixed angle, which is then analyzed with an autocorrelator. The resulting correlation function has the particle diffusion coefficient as one of its arguments, which can be used to calculate the hydrodynamic radius of the particle.

2.2.2. Small-Angle X-Ray Scattering

Small-angle X-ray scattering (SAXS) is another solution scattering technique that has gained prominence in recent years. Although it is not suitable for structural analysis at the atomic level, SAXS often provides valuable structural information at lower resolution (overall shape of the protein, as well as its quaternary and tertiary structure). The molecular size range amenable to SAXS spans the range 10–1000 Å, thus providing a means to characterize not only single proteins, but also very large macromolecular assemblies (15). As the name infers, SAXS measures an interference pattern of elastically scattered X-rays, which is usually expressed as a function of a scattering vector length Q:

$$\mathrm{Q} = |\vec{\mathrm{Q}}| = \frac{4\pi}{\lambda}\sin\theta \qquad (2\text{-}2\text{-}3)$$

where λ is the wavelength of the X-ray beam and 2θ is the scattering angle. The measured scattering profile $I(\mathrm{Q})$ is a contrast between the X-rays scattered from the protein molecules in solution and the equivalent volume of solvent displaced by them (15):

$$I(\mathrm{Q}) = \left\langle \left| \int d^3\vec{\mathrm{r}}(\rho(\vec{\mathrm{r}}) - \rho_{\mathrm{s}}(\vec{\mathrm{r}}))\mathrm{e}^{-i\vec{\mathrm{Q}}\cdot\vec{\mathrm{r}}} \right|^2 \right\rangle \qquad (2\text{-}2\text{-}4)$$

where the averaging is done with respect to all orientations of vector $\vec{\mathrm{Q}}$. The lack of translational order (lattice) within the protein solution leads to continuous distribution of $I(\mathrm{Q})$ (unlike distinct Bragg reflections arising due to the translationally ordered molecular structure in crystals). The contrast factor $\Delta\rho$ [difference between the electron densities of protein molecules (ρ) and solvent (ρ_{s})] is typically on the order of $0.12\,\mathrm{e}^{-}/\mathrm{Å}^3$ (four times lower that ρ alone). As a result, protein scattering strength in aqueous solution is only $\sim$10% of what it would be in a solvent-free environment (16). More rigorous treatment of the scattering profile $I(\mathrm{Q})$ accounts for scattering from the protein hydration shell, whose electron density, as well as other physical properties, may be different from that of the bulk solvent.

One of the most common ways to interpret the SAXS data employs the inverse Fourier transform (FT) of $I(\mathrm{Q})$ to yield the radial Patterson distribution:

$$P(r) = \frac{1}{2\pi^2}\int [I(\mathrm{Q})\cdot \mathrm{Q}r\cdot\sin(\mathrm{Q}r)]d\mathrm{Q} \qquad (2\text{-}2\text{-}5)$$

The Patterson distribution is related to the probability distribution of the distances between the scattering centers (atoms) within the scattering particle. As such, it provides information on the overall shape of the scattering particle and its dimension (15). The Patterson distribution can also be used to calculate the gyration radius of the scattering particle:

$$R_g = \frac{\sqrt{\int P(r)\cdot r^2 dr}}{\sqrt{\int P(r)dr}} \qquad (2\text{-}2\text{-}6)$$

a parameter that reflects both its size and shape. Recently, extensive efforts have been directed toward developing methods that restore low-resolution shape, as well as the internal structure of macromolecules in solution based on isotropic scattering data (17). All of the above considerations relate to a homogeneous solution of nearly identical (although randomly oriented) scattering particles. Sample heterogeneity (either the presence of multiple protein species or conformational inhomogeneity) poses a serious challenge to SAXS. In this case, a measured scattering profile would contain contributions from all components, weighted by their fractional concentrations. This difficulty can be overcome in some cases by using singular value decomposition (SVD) to characterize individual protein conformers (18) or protein aggregates (19).

2.2.3. Cryo-Electron Microscopy

Cryo-electron microscopy (cryo-EM) is an imaging technique that often allows large macromolecular complexes to be visualized at low resolution. This technique exploits the fact that images of macromolecules represent their 2D projections. Three-dimensional structural data can be obtained from such projections if the data from "different views" are combined and processed using various reconstruction techniques (20). In a few cases it can actually provide structural information at a resolution high enough to derive atomic coordinates, although typically it is not possible to determine macromolecular structure at a resolution >10–30 Å (21). Combination of the low-resolution cryo-EM data on macromolecular assemblies and high-resolution and X-ray crystallographic structures of their constituents can provide structure of large sub-cellular assemblies at near-atomic resolution by "docking" the high-resolution structures into the low-resolution maps (22).

The energy of electrons used in electron microscopes corresponds to a wavelength range of 0.0015–0.040 Å. The scattering power of the electrons is significantly higher than that of X-rays, hence there are much smaller sample requirements. An important distinction (and a serious drawback) of electron microscopy from other scattering methods is the necessity of a high vacuum environment in the microscope. This may be achieved by rapid freezing of the sample and maintaining it at or below liquid nitrogen

temperature to minimize solvent sublimation during the experiment. Freezing also reduces radiative damage to the macromolecular structures, although even such samples typically do not tolerate >10–15 electrons/Å^2 (20). Therefore, in order to obtain cryo-EM data at a high signal-to-noise ratio (S/N), 10,000–100,000 individual particle images need to be aligned and averaged (20). The existence of any kind of symmetry in the sample enhances the S/N in the cryo-EM images, although imaging of asymmetric randomly oriented single particles is also possible. In fact, it has been possible in some cases to obtain structural data on single molecules (provided they exceed the lower size limit of several hundred kDa) (23). Cryo-electron microscopy provides a means of capturing different conformational states of macromolecular assemblies by imaging the complexes that are "trapped" at different stages of their interactions and conformational changes (23,24).

Methods of cryo-EM analysis also can be used to characterize biological macromolecules and their assemblies within the cells. Utilization of this technique (usually referred to as cryo-electron tomography, or cryo-ET) to study *in vivo* processes is based on the precept of cryogenic preservation of the native (fully hydrated) biological structure (25). Several examples of using cryo-ET to characterize structure of macromolecular complexes within their cellular context demonstrate promise to reveal the internal structure of organelles and indeed entire cells at molecular resolution (25–28). Resolutions of 5-8 nm have already been achieved and the prospects for further improvement are good (29). Since many intracellular activities are conducted by complexes in the megadalton molecular weight range with dimensions of 20–50 nm, current resolutions should suffice to identify many of them in tomograms, although the residual noise and dense packing of cellular constituents hamper interpretation (29). Medalia et al. (30) relatively recently collected cryo-ET data on vitrified eukaryotic cells.

2.2.4. Neutron Scattering

Small angle neutron scattering (SANS) is an emerging technology (as far as applications to biological systems). It utilizes the same principle as SAXS, although the scattering centers are predominantly nuclei (as opposed to electrons in SAXS). Analysis of the SANS profiles can reveal important information related to the mass and the overall shape of the scattering particles, as well as the distribution of interatomic vectors within a molecule (31). Unfortunately, such experiments remain a rarity, as access to the research facilities possessing neutron beam generators remains limited.

2.3. NMR SPECTROSCOPY

Nuclear magnetic resonance (NMR) spectroscopy has for many years been at the forefront of protein structure determination. Spin energy levels of nuclei, when placed in a magnetic field, become split by the Zeeman effect. Upon application of radio frequency (rf) radiation, each nucleus within a molecule can be caused to resonate at a specific frequency that is highly dependent on its local environment. In the case of proteins, the naturally abundant proton (^{1}H) has nuclear spin $I = \frac{1}{2}$, and by isotopic enrichment methods, other spin-$\frac{1}{2}$ nuclei ^{13}C and ^{15}N can be relatively easily incorporated into expressed proteins. A battery of complex heteronuclear pulse sequences have been developed along with multidimensional NMR to achieve high-resolution structure determination for relatively small proteins. Liquid helium supercooled magnets are now commercially available with magnetic field strengths upward of 20 T (900 MHz ^{1}H frequency), enabling high-resolution spectroscopy of quite large biomolecules.

In a magnetic field B_0, the spin states of a nucleus become split into equally spaced energy levels according to the magnetic field strength and the gyromagnetic ratio γ, a constant for any given nucleus:

$$E_m = -\frac{\gamma m h B_0}{2\pi} \tag{2-3-1}$$

where h is Planck's constant. For a spin-$\frac{1}{2}$ nucleus there are two possible values of the magnetic quantum number $m = (-\frac{1}{2}, +\frac{1}{2})$, giving rise to two Zeeman nuclear energy levels E_m under the influence of B_0. Nuclear transitions between the two energy levels may be effected by application of electromagnetic radiation. The difference in energy levels:

$$\Delta E_m = \frac{\gamma h B_0}{2\pi} \tag{2-3-2}$$

corresponds to radiation of frequency ν_0:

$$\nu_0 = \frac{\omega_0}{2\pi} = \frac{\gamma B_0}{2\pi} \tag{2-3-3}$$

Thus, as in other spectroscopic methods, transitions between the nuclear energy levels may be produced by irradiation. In older NMR spectrometers, the rf spectrum was scanned and each individual resonance could be detected as absorbance. However, whereas in optical spectroscopy irradiation is with a constant light source and absorptive or fluorescent properties are measured, in the case of modern pulsed NMR the highest energy spin level is preferentially populated by irradiation and then the relaxation to the ground state is measured. Upon inversion of nuclear spin by application of an rf electromagnetic pulse, the magnetization relaxes back to the ground state, and in doing so precesses around the B_0 magnetic vector at its characteristic Larmor precession frequency ω_0. It is this resonance frequency that is detected by the rf receiver coil of an NMR spectrometer.

The power of NMR for structural characterization comes from the fact that each nucleus within a molecule will have a slightly different Larmor frequency, depending on its local environment, that is, shielding by local magnetic fields of nearby electronic motion. Thus an NMR spectrum will contain peaks corresponding to resonances from each nucleus in the sample, each with a frequency (known as a chemical shift, δ) characteristic of the chemical environment. Rather than using absolute frequency units, by convention the resonant frequencies are denoted as parts per million (ppm) differences, relative to an internal standard, of the carrier frequency (for an 11.7-T magnet, for ^{1}H nuclei $\nu_0 \sim 500$ MHz). The so-called chemical shift scale (δ, measured in ppm; $\delta = \omega_0/\nu_0$) is then independent of the magnetic field strength B_0, making data acquired on different instrumentation more readily comparable. For protons, the chemical shift values generally fall in the range 0–10 ppm (relative to an internal reference compound, e.g., tetramethylsilane defined as $\delta = 0$ ppm).

In modern instrumentation, generally all protons within a molecule are simultaneously excited by a broadband rf pulse. This causes all nuclei to simultaneously resonate at their characteristic frequencies, which can be detected as a time domain signal image current by the receiver coil. This signal can then be separated into individual signals in the frequency domain by application of a FT:

$$S(\nu) = \mathrm{FT}\{s(t)\} = \int_{-\infty}^{\infty} s(t)\mathrm{e}^{-i2\pi\nu t}dt \qquad (2\text{-}3\text{-}4)$$

where $s(t)$ and $S(\nu)$ are the time and frequency domain signals, respectively. The final NMR spectrum consists of frequency domain data with peaks spread across the chemical shift scale representing each resonance. In addition, the area under each peak is proportional to the number of nuclei resonating at that frequency.

An important factor, one that makes NMR an incredibly powerful structural tool, is the "cross-talk" between nuclei within a molecule. Perturbation of local electron clouds gives rise to a scalar coupling between nuclei that are covalently bonded and is significant for nuclei separated by up to four bonds. The effect of coupling is to alter the energy levels of the two spin states and produce new energy levels depending on whether the spins of the two nuclei are aligned in parallel or antiparallel directions. This leads to multiplet fine structure in the observed NMR peaks. The number and nature of adjacent protons gives rise to specific coupling patterns and coupling constants (differences in energy levels, denoted J and measured in Hz), which can be used to assign covalent structure. In proteins, each amino acid side chain gives a characteristic scalar coupling pattern that is invaluable for spin system assignment.

One of the most important NMR phenomena for protein structure determination is an alternative coupling scheme that operates via "through-space" interactions. Once a spin has been inverted by an rf pulse it relaxes to the ground state via various mechanisms. If other nuclei are close enough in space (< 5 Å for ^{1}H), then a dipolar coupling known as the nuclear Overhauser effect (NOE) can also occur. The proximity of other nuclei allows alternative relaxation pathways for the spin, so this manifests itself as a change in the time taken for a nucleus to relax to its ground state. The relaxation effect falls off rapidly with internuclear distance (r) following the relationship NOE $\propto 1/r^6$, hence the NOE intensity can be measured and used to provide distance constraints in structural calculations.

In the simplest NOE experiment, the resonance of a single nucleus is saturated by continuous radiation (this effectively equalizes the population of the two spin states), followed by broadband excitation of all other nuclei, and subsequent detection. Nuclei that are sufficiently close in space to the saturated nucleus will experience a signal enhancement. The resultant spectrum can be subtracted from a normal spectrum obtained without saturation, yielding a difference spectrum, which will contain peaks only from nuclei that are spatially close to the target nucleus. The more complex two-dimensional NOE spectroscopy (2D NOESY) experiment correlates all nuclei that are in proximity to one another and, once Fourier transformed, produces a spectrum from which cross-peak intensities can be directly correlated to maximal distances. Using the known allowable backbone torsion angles with the network of distance constraints that may be obtained from NOE experiments allows structural models to be built and refined for proteins.

As the number of nuclei in a molecule increases, the limited chemical shift range accessible (generally -2 to 12 ppm for protons), leads to a very crowded NMR spectrum. One way to alleviate this problem is to add a variable delay, or evolution time, into the pulse sequence. In subsequent experiments, this evolution time is increased by a fixed amount each time, leading to a 2D array. Fourier transformation of the array in both dimensions produces a 2D spectrum in which the diagonal corresponds to a simple NMR spectrum, but now cross-peaks are observed that, depending on the chosen pulse sequence, correlate resonances that are coupled either in a scalar or dipolar manner to one another. This greatly simplifies the process of spectral assignment, since it is now a relatively much simpler matter to work out which nuclei are correlated, either through bonds or through space. Multidimensional NMR experiments have evolved complex descriptive acronyms (e.g., COSY correlation spectroscopy) to describe scalar couplings and NOESY (for dipolar couplings). In the latter case, the intensity of the cross-peak is related to the relaxation

effect of the neighboring nucleus, and can thus be used as a distance constraint for structure determination.

2.3.1. Heteronuclear NMR

In naturally occurring proteins, the only spin-$\frac{1}{2}$ nucleus present at high abundance is ^{1}H. With modern molecular biology techniques, however, isotopically enriched media can be used quite readily to incorporate ^{15}N and ^{13}C into proteins at high levels. In the case of proteins expressed in *Escherichia coli* (*E. coli*), cells are grown on a minimal medium containing $^{15}NH_4Cl$ and ^{13}C-glucose as the sole sources of nitrogen and carbon, respectively. This generally produces reasonable yields of isotopically labeled protein. Now, in the NMR spectrometer, heteronuclear, as well as homonuclear, coupling come into play. Protons will experience scalar coupling not only from neighboring protons, but also from ^{13}C and ^{15}N nuclei. Thus we can take advantage of the chemical shift dispersion not only of protons, but also the other nuclei, and complex pulse sequences have been developed, spanning multiple dimensions, to trace backbone resonance assignments and provide further constraints for structure refinement of biomolecules (Fig. 2.1).

To date, the upper limit for complete resonance assignment of proteins remains at ~30 kDa, due to resonance overlap and unfavorable relaxation times of large biomolecules. Further technological advancements both in magnetic field strength and probe design promise to increase that limit. The development of the transverse relaxation optimized spectroscopy (TROSY) pulse sequence (32) also has allowed a major step forward by removing the line-broadening effect of slowly tumbling molecules in solution. In addition, the partial alignment of proteins within lipid bicelles shows much promise (33,34).

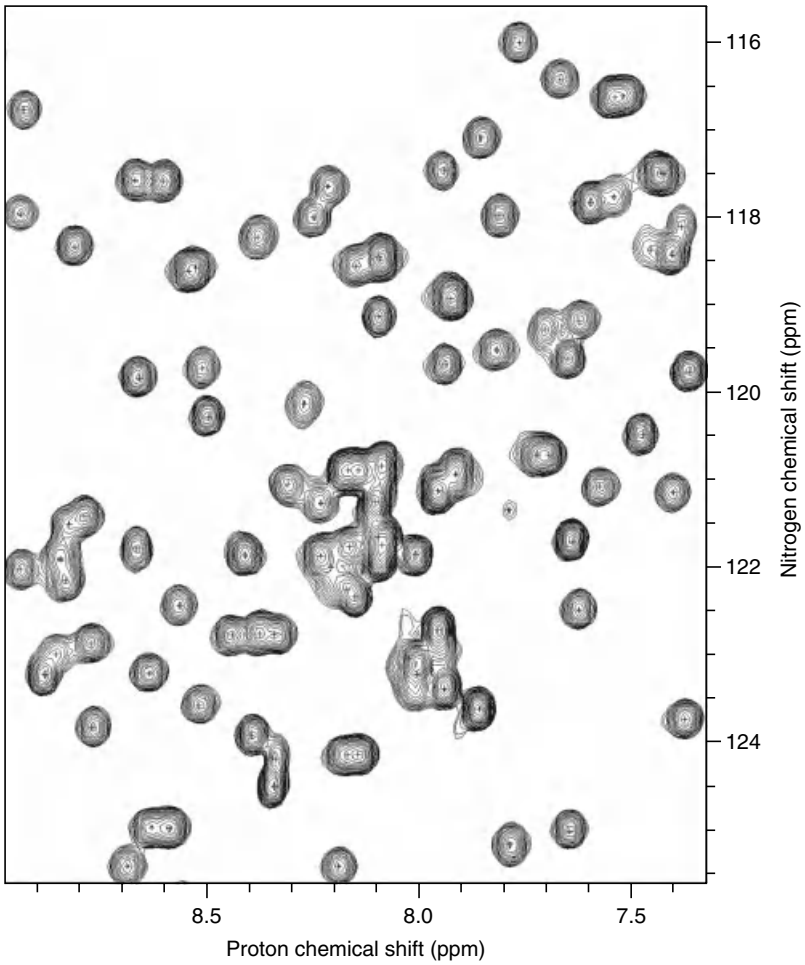


Figure 2.1. Amide *fingerprint* region of the $^{15}N-^{1}H$ HSQC spectrum of a protein. Each contoured peak represents an amide proton whose ^{1}H resonant frequency and that of the adjacent ^{15}N are correlated. In a folded protein, these resonances are well resolved due to the chemical shifts induced by persistent structure. [Figure courtesy of Dr. Joanna Swain, University of Massachusetts-Amherst.]

In addition to the experiments that seek to elucidate structure, there are NMR strategies aimed specifically at measuring protein dynamics. As mentioned above, relaxation of nuclei after an rf pulse can be affected by a number of factors, one of which is intramolecular internal motions. The ^{15}N nuclei, most notably backbone amides, are particularly susceptible to dynamics-related relaxation via the NOE. Measurement of the ^{15}N NOE and relaxation times at each amide along the backbone of a protein allows extraction of order parameters, which can be used to determine the dynamic nature of individual segments within a protein. This has enabled flexible regions within proteins to be identified, with correlation times in the picosecond-to-microsecond time range, and can be correlated to intermolecular interactions and functional dynamics.

2.3.2. Hydrogen Exchange by NMR

As described further in Section 4.4 and Chapter 5, the rate at which labile backbone amide protons exchange with bulk solvent also serves as a measure, not only of surface accessibility, but also of dynamics. Amides that are normally buried or else hydrogen bonded, can become exposed to solvent during dynamic opening (or local unfolding) events. If the bulk solvent is heavy water (^2H_2O or D_2O), then an exchange reaction $>$NH $\rightarrow >$ND will cause the disappearance of the corresponding peak in the NMR spectrum, since deuterium is an NMR-invisible isotope (Fig. 2.2). Thus, measuring intensities of the amide resonances as a function of time in deuterated solvent can be a powerful residue-specific monitor of dynamics (35,36). Experimental conditions, such as temperature and pH, can be manipulated to extract thermodynamic details, and to obtain actual measures of the free energies involved in certain dynamic events (37–42).

Likewise, during the folding of a protein, amides that form stable structure earliest in the folding process will become protected against exchange with solvent. In combination with a quenched-flow apparatus, pulse-labeling experiments can be performed that can be used to monitor protein folding on the millisecond time scale. Unfolded protein is allowed to refold for a variable time before labeling all exposed amides with deuterium. For the purposes of studying hydrogen exchange, the COSY (correlation of amide NH and C_αH protons) and ^{15}N-heteronuclear single quantum coherence spectroscopy (^{15}N-HSQC) which correlates amide protons with the attached nitrogen in uniformly ^{15}N-labeled proteins, experiments are most useful. In a fully protonated sample, these correlation peaks will have maximum intensity, whereas if the proton becomes replaced by a deuteron due to an exchange event, then it will no longer be visible [deuterium (D) nuclei are not observable in ^{1}H NMR experiments]. Thus, in a refolding experiment where unprotected amides are labeled with D, the only peaks observed are those that correspond to residues protected against exchange before the isotope-labeling event. Intensities of amide resonances can then be used to determine the order of formation of structural regions during folding (43–45).

A natural extension of this, which has yielded some exciting results, is the ability to use a flow system into the NMR spectrometer and effectively perform the whole

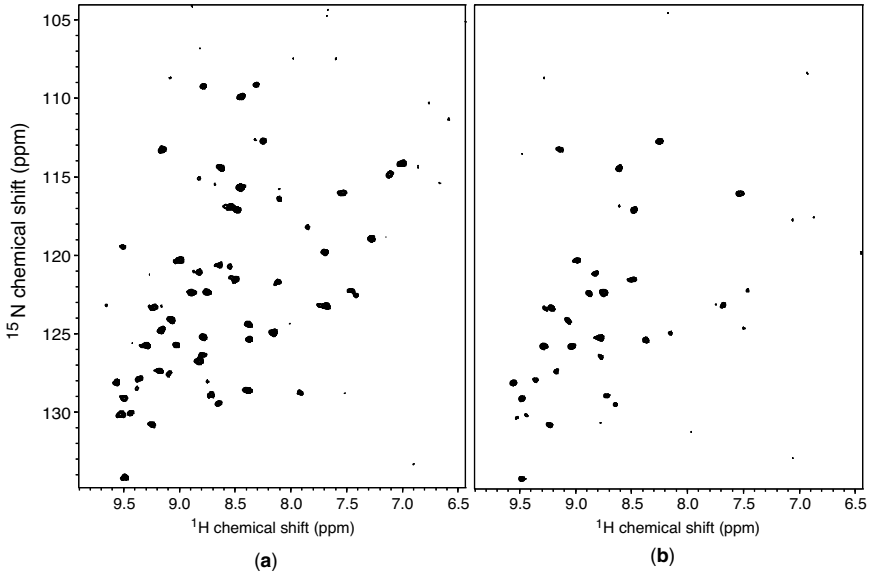


Figure 2.2. Dynamics of the C-terminal domain of DnaK in D_2O measured by HDX NMR at (*a*) $t = 0$, (*b*) $t = 24$ h. The NMR resonances observed at the earlier time point disappear over time as ^{1}H is replaced by ^{2}H. [Figure courtesy of Dr. Joanna Swain, University of Massachusetts-Amherst.]

experiment online. Stopped-flow NMR has been used to investigate the folding of proteins in real-time (46,47), although the time resolution is somewhat limited compared with stopped-flow optical methods. As instrument sensitivities are improved and higher throughput is required, various hybrid instruments have appeared, notably a combination LC system with an outlet to flow-through NMR and electrospray mass spectrometry.

The NMR method, while clearly extremely useful and high resolution, is limited by the need for complete assignment of all resonances. For large proteins, resonance overlap and peak broadening become significant issues. Additionally, other factors (e.g., paramagnetic ligands) may also preclude the use of NMR methods. However, perhaps one of the more serious limitations of NMR to measure protection in pulse-labeling experiments is that it can only measure protection at amino acid residues that are sufficiently protected against exchange during the time course of sample workup and of the NMR experiment itself. Folding experiments *in vitro* are normally performed at low concentrations ($<10\,\mu M$) to avoid the possibility of aggregation or other nonphysiologically relevant misfolding events. However, for NMR experiments, samples in the millimolar concentration range are required, so time consuming concentration steps are often required. These stages may enable further undesirable hydrogen–deuterium exchange (HDX) processes to occur. Even with fast NMR experiments the total time for sample preparation and measurement may be several hours. This workup time could potentially lead to loss of information in regions of the protein that are less highly protected in the native structure.

The field of NMR spectroscopy is far too wide-ranging to cover anything more than the basics in this text. We refer the reader to a number of excellent books on the subject (48,49).

2.4. OTHER SPECTROSCOPIC TECHNIQUES

2.4.1. Cumulative Measurements of Higher Order Structure: Circular Dichroism

Circular dichroism (CD) spectroscopy utilizes differential absorption of left- and right-polarized light (usually within the 170–700-nm range) by chiral molecules. Since most biological molecules possess multiple chiral chromophores, their CD spectra may provide important insights on the 3D arrangement of such macromolecules with varying specificity. Specifically, CD spectroscopy is often used to measure the overall secondary structural content of peptides and proteins [the far-ultraviolet (UV) region, 190–250 nm], chirality due to the unique arrangement of the aromatic side chains and/or disulfide bridges within the protein (250–300 nm), as well as protein interaction with chiral ligands and cofactors (e.g., Soret band). Although the assessment of protein conformations afforded by CD spectroscopy typically yields only a low-resolution picture, the experiments are relatively simple and not particularly demanding in terms of sample work-up and consumption. As a result, CD measurements are often used to complement measurements by other techniques.

Far-UV CD: Evaluation of Secondary Structural Content of Peptides and Proteins. The most ubiquitous chromophore in peptides and proteins is the peptide amide bond, which gives rise to two characteristic absorption bands in the far-UV region: a $\pi \rightarrow \pi^*$ transition at 190 nm and a weaker (and broader) $n \rightarrow \pi^*$ transition at 210 nm (50). Different forms of secondary structure arrange the peptide bonds in a very specific nonrandom asymmetric fashion, giving rise to characteristic and recognizable CD spectra in the far-UV region. Thus, spectra of proteins with mostly α-helical content ($\alpha\alpha$ proteins) have pronounced negative bands at 222 and 208 nm (Fig. 2.3), a feature that is notably absent from the spectra of proteins rich in β-sheets. The latter are usually characterized by a single minimum whose position may vary within the 210–230-nm range, depending on the protein (Fig. 2.3). Far-UV CD spectra of some β-rich proteins may actually resemble model spectra of unordered (random coil) polypeptides, exhibiting a minimum below 200 nm (51). Proteins that contain both α-helices and β-strands have contributions from both secondary structural elements in their CD spectra, however, it is very difficult to distinguish proteins with separate ($\alpha + \beta$) or intermixed (α/β) regions based solely on CD data. These two classes are often treated as a combined $\alpha\beta$ class for the purposes of CD analysis (52).

There are numerous algorithms that can be used to calculate the secondary structural content of a given polypeptide or a protein based on its far-UV CD spectrum. The most straightforward way to obtain a reasonable estimate of the helical content (f_h) uses mean residue ellipticity [θ] at 208 or 222 nm (53):

$$f_h = -\frac{[\theta]_{208} + 4{,}000}{29{,}000} \qquad (2\text{-}4\text{-}1)$$

$$f_h = -\frac{[\theta]_{222} + 3{,}000}{33{,}000} \qquad (2\text{-}4\text{-}2)$$

More sophisticated treatments of this problem take into account the finite length of helical segments within the protein (both $[\theta]_{208}$ and $[\theta]_{222}$ diminish as helical segments become shorter).

The secondary structural content of the $\alpha\beta$ proteins can also be estimated using various empirical methods. An underlying assumption is that the CD spectrum can be presented as a linear combination of individual secondary

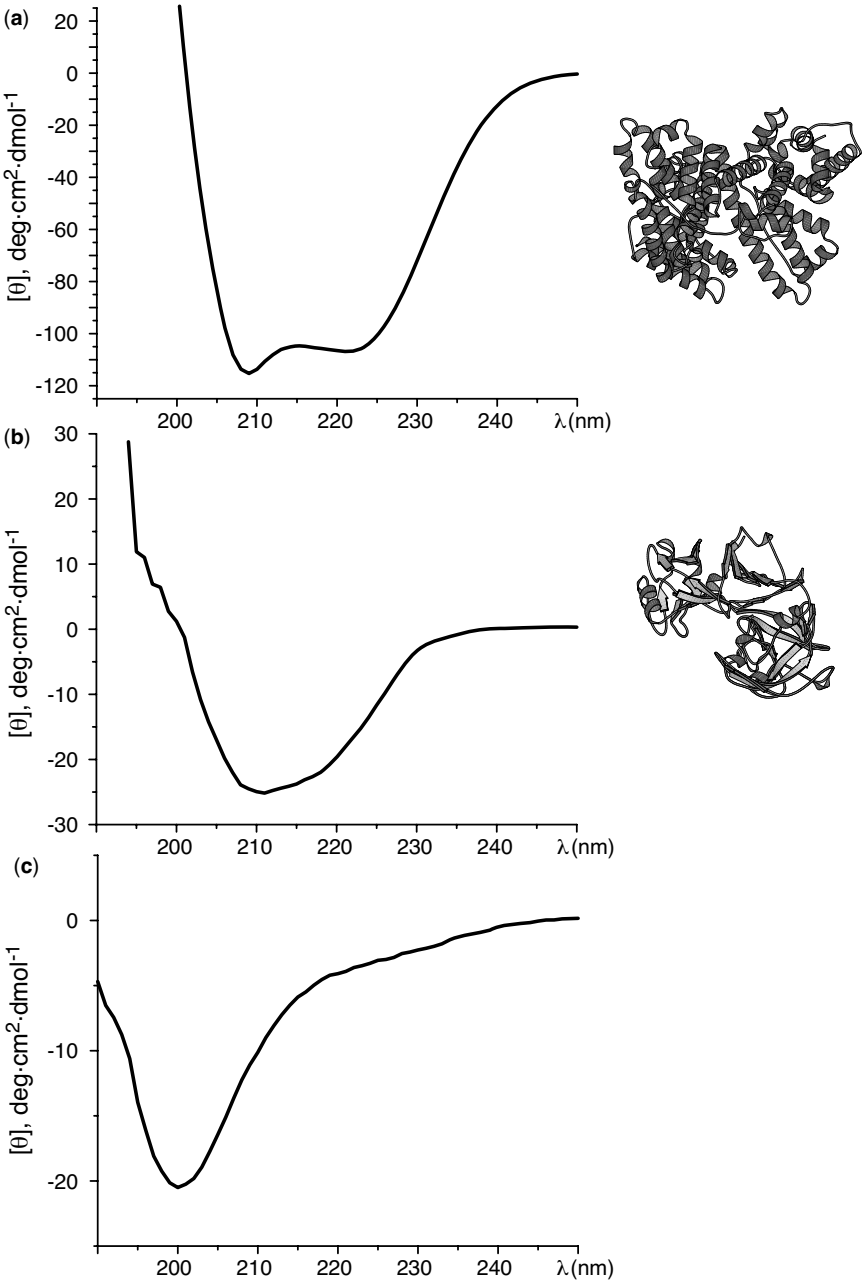


Figure 2.3. Far-UV CD spectra of an α-helical protein bovine serum albumin (BSA) (*a*), a mostly β-sheet protein β-lactoglobulin (*b*), and a mostly disordered protein β-casein (*c*).

structural components, including random coil conformation (more rigorous treatment also takes into account contributions from aromatic chromophores and noise). Usually, CD spectra of a set of reference proteins with known structures are used to determine the component secondary structure spectra. Several algorithms have been designed for these purposes, and an interested reader is referred to several excellent recent reviews on the subject (50).

Near-UV CD: Interactions of Aromatic Side Chains and Disulfide Bonds. The near-UV absorption spectra of proteins are dominated by the bands of aromatic side chains (Phe, Tyr, and Trp), and their unique arrangement within the protein structure usually gives rise to CD signal (Figs. 2.4 and 2.5). The intensity of such peaks depends on the number of aromatic amino acid residues in the protein, as well as on flexibility of the protein. The environment of the aromatic groups and protonation state (for Tyr) also influence the

signal. A disulfide bond is another intrinsic protein chromophore that contributes to the near-UV CD spectra (near 260 nm). However, this band is usually rather weak and broad, making it difficult to distinguish the disulfide signal from that of the aromatic residues. If two (or more) aromatic residues are located in close proximity, coupling of the electronic transitions usually occurs, leading to significant enhancement of the CD signal. These local interactions are more important than the sheer number of the aromatic residues in determining the appearance of the near-UV CD spectra (54). In fact, in many cases it is possible to use near-UV CD spectroscopy to detect local conformational changes that affect the arrangement (and, therefore, transition coupling) of such proximally clustered aromatic side chains.

Due to their complexity and the multiplicity of factors influencing the appearance of the near-UV CD spectra, the latter cannot be interpreted as easily as the far-UV CD spectra. Nonetheless, the near-UV CD measurements can also be very useful, particularly in a situation when only a qualitative assessment of protein ternary structure (or lack thereof) is required. For example, near-UV CD spectroscopy is often used as a means to monitor progressive loss (or gain) of protein tertiary structure in response to changing solvent conditions (Fig. 2.4) or point mutations. Still, even such qualitative analyses need to be carried out with great care, as the appearance of the protein near-UV CD spectrum can be influenced by a variety of extrinsic factors.

CD Signals Due to Extrinsic Chromophores. Proteins themselves (i.e., polypeptide chains without bound ligands) usually do not give rise to a CD signal >300 nm. In many cases, however, this region does contain CD bands that are collectively referred to as extrinsic CD bands (50). These bands represent transitions in ligands, which may have inherent optical activity (due to chirality) or that become optically active due to the coupling of electronic transitions on the protein chromophores and the extrinsic chromophore, as well as mixing of transitions of differing symmetry (50). In the latter two cases, the extrinsic chromophore does not have to be optically active to produce a CD signal, which thus provides valuable information on the protein–ligand interaction.

One of the specific examples of CD spectroscopic studies of protein–ligand interaction that will be used in later chapters is the heme–polypeptide interaction in hemoproteins. The prosthetic heme group has an absorption band at 410 nm (Soret band), which remains "CD silent" unless the heme is bound to a protein. Asymmetry introduced by positioning of the heme in a binding pocket within the protein leads to appearance of an abundant signal in the CD spectrum of the protein (Fig. 2.4). Both position and intensity of this band are influenced by the heme environment within the binding pocket, and are often used as

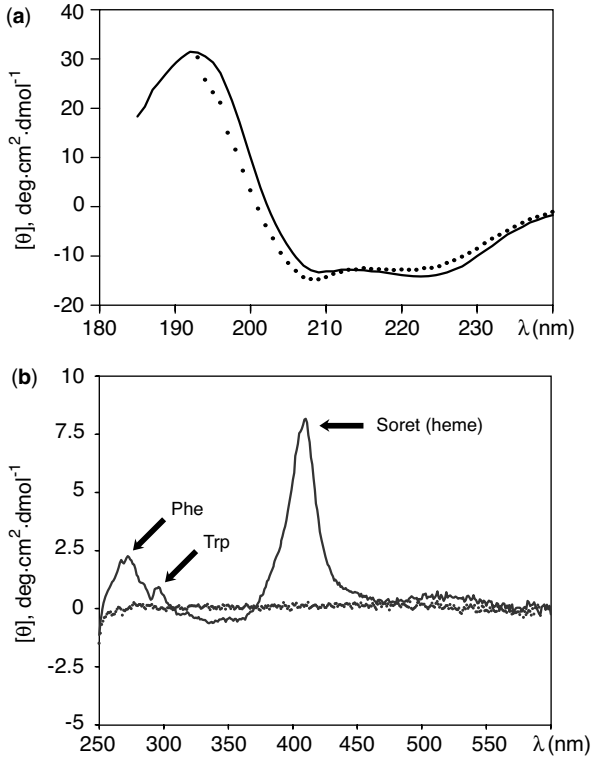


Figure 2.4. Acid unfolding of myoglobin monitored with CD spectroscopy in the far-UV (*a*) and near-UV/vis (*b*) regions. The solid lines represent CD spectra of myoglobin acquired under the near-native conditions (10 mM ammonium acetate); the dotted lines represent the spectra acquired in H_2O/CH_3OH (water/methanol) solution acidified with 1% acetic acid (CH_3COOH) (by volume). Disappearance of the Soret band indicates heme dissociation from the protein, and the disappearance of the aromatic side-chain bands (Phe, Tyr, Trp) indicates loss of tertiary contacts, while a significant proportion of the secondary structure (α-helical) is retained.

reporters of the protein–heme interaction. Figure 2.4 illustrates how different regions of the CD spectra of myoglobin can be used to characterize higher order structure of this protein at different levels (near-UV, secondary structure; far-UV, tertiary fold; and Soret band, heme–protein interaction).

It should probably be emphasized once again that significant changes in the near-UV CD spectra do not necessarily indicate large-scale protein conformational changes. Numerous extrinsic factors, as well as subtle changes in the environment of aromatic side chains, often lead to profound alterations of certain CD bands. An example of such behavior is shown in Figure 2.5, where replacement of two metal ions within a 78 kDa protein transferrin results in profound changes in the near-UV CD profiles. However, the protein tertiary structure is not affected by the Fe^{3+} to Ga^{3+} (or In^{3+}) substitution. Rather, these rather dramatic changes

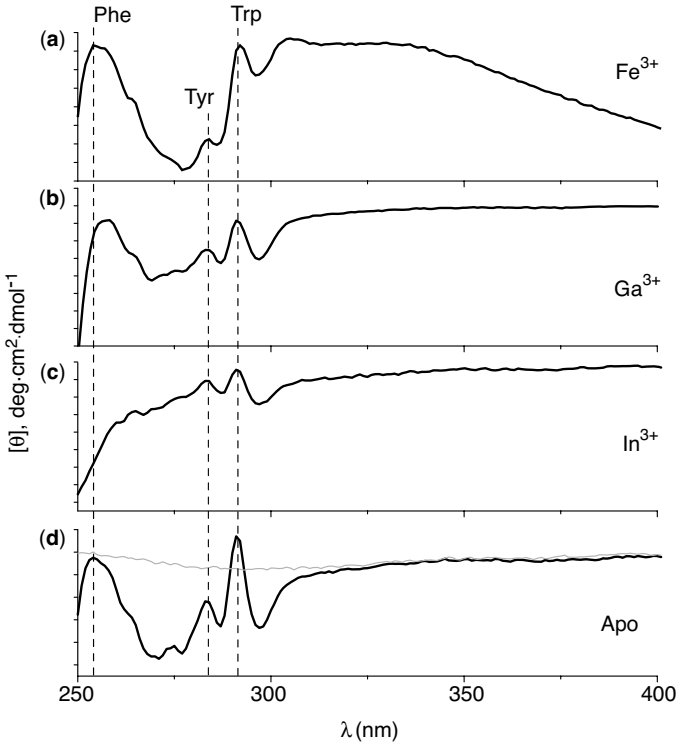


Figure 2.5. Near-UV CD spectra of human serum transferrin substituted with Fe^{3+} (*a*), Ga^{3+} (*b*), In^{3+} (*c*), and the apo-form of the protein (*d*). [Spectra are courtesy of Dr. Mingxuan Zhang, presently at Biogen IDEC, Cambridge, MA.]

in the appearance of the CD spectra reflect changes in the coupling between the electronic transitions of the metal ions and those of the aromatic side chains. Another example is evolution of the Soret band in the CD spectrum of myoglobin at a pH above neutral. Although the heme group remains bound to the protein within this pH range, there is a noticeable red shift and a gradual demise of the Soret band as the solution pH is increased >8. This shift is caused by protonation of some of the residues in the heme-binding pocket of the protein, a process that does not compromise the structural integrity of the heme–protein complex. These two examples emphasize great care that must be exercised when interpreting both near-UV and visible region CD data.

CD Spectroscopy of Other Biopolymers: Oligonucleotides. The nucleic acid chromophores are the bases, which have rich high-intensity absorption spectra in the region <300 nm. The bases do not have intrinsic optical activity; however, they are attached to asymmetric sugars (deoxyribose or ribose rings), which induce CD bands in the absorption region of nucleic acids. Oligonucleotide base stacking in solution forces the entire chain to assume a helical structure, resulting in "super-asymmetry". Electronic transitions in the neighboring bases interact with each other, leading to CD signal amplification. Formation of double-stranded helices by complementary oligonucleotides also leads to super-asymmetry, hence high-intensity and information-rich CD spectra. Due to its sensitivity to the base–base interactions, CD spectroscopy is often used to monitor changes in the secondary structure of oligonucleotides (55). CD spectroscopy has also been applied to study nonclassical nucleic acid structures (triplex, quadruplex, parallel DNA, etc.). A detailed review of such applications can be found in (56).

2.4.2. Vibrational Spectroscopy

Vibrational Absorption (Infrared, IR) Spectroscopy. This is another popular tool to study protein higher order structure and dynamics. Because of the enormous number of normal modes of vibration in a typical protein, the vibrational bands overlap, giving rise to very complex IR spectra. Nevertheless, it is often possible to extract valuable information from such spectra that provide valuable insight into protein behavior. The basic principles of protein vibrational spectroscopy can be understood by considering a simple harmonic oscillator, whose frequency is calculated as:

$$\nu = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k(m_1 + m_2)}{m_1 m_2}} \qquad (2\text{-}4\text{-}3)$$

where m_1 and m_2 are the masses of the two atoms in the oscillator and k is a force constant. Any change in the force constant (determined by the interatomic bond electron density) will result in a frequency change. The intensity of the absorption band is determined by a number of factors. For an ideal harmonic oscillator, the transition can occur only to the next vibrational level. In the mid-IR region, the majority of oscillators are not excited at room temperature; hence, only transitions to the first excited state are observed. The probability of such transitions will be determined by the square of the transition dipole moment (TDM) (57):

$$\mathrm{TDM} = \left\langle \frac{\partial \mu}{\partial R(R_0)} \right\rangle \sqrt{\frac{h(m_1 + m_2)}{8\pi^2 \nu m_1 m_2}} \qquad (2\text{-}4\text{-}4)$$

where μ is the dipole moment and h is Planck's constant. The $\langle \partial\mu/\partial R(R_0) \rangle$ term represents the electronic contribution to the transition probability (by linking the latter to the change of the dipole moment at the equilibrium position R_0). Obviously, stronger absorption is favored by larger dipole changes, which are often correlated with significant bond polarities. Therefore, strong absorption bands are often observed for polar bonds (e.g., carbonyl), while carbon–carbon bands are typically much weaker or absent from the vibrational spectra. The following brief discussion is intended to provide a brief overview of some of the key features of the vibrational spectra of proteins.

Amide Group Absorption. There are several bands associated with the vibrations of the protein amide groups. The so-called *amide I* band (near 1650 cm^{-1}) is most commonly used for secondary structure analysis, as it is affected by protein backbone structure (57). This band is mostly due to the carbonyl stretching with minor contributions from other vibrations (out-of-phase C−N stretching, C−C−N deformation, and N−H bending). Although both theoretical considerations, as well as experimental work carried out with model proteins, suggest that different secondary structural elements give rise to distinct amide I bands, they usually overlap in the absorption spectra. As a result, a broad amide I band is typically observed for a majority of proteins (Fig. 2.6). Two different approaches are often used to analyze protein secondary structure based on the appearance of the amide I band. The first aims at decomposing the amide I band into "component bands" corresponding to different types of secondary structure. This approach employs various mathematical procedures to resolve the component bands (by using Fourier deconvolution, taking second derivative, etc.), followed by fitting the experimental band with a sum of components, each of which is then assigned to a certain secondary structural element (57). The second approach utilizes a "calibration set of spectra" (i.e., spectra of proteins with known structure) to perform pattern-recognition calculations. A chemometric procedure (factor analysis) is then used to reduce a large number of spectra in the calibration set to a smaller set of a few linearly independent basis spectra. This set is used to reconstruct the "unknown" spectrum.

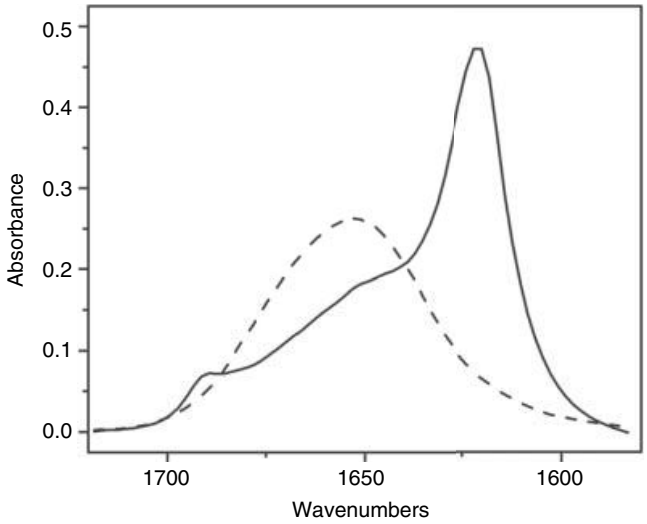


Figure 2.6. Fourier transform infrared (FTIR) spectrum of a synthetic prion peptide at 25°C (solid line) and 75°C (dashed line). The peptide is predominantly in *a β-sheet* conformation at low temperatures, as evidenced by the two sharp peaks at 1621 and 1691 cm^{-1} in the scan. The beta sheets then melt as temperature increases, so that the spectrum at high temperature shows a single broad peak at 1653 cm^{-1}, characteristic of a random coil. [Figure courtesy of Dr. Sarah Petty (presently at College of the Holy Cross).]

The *amide II* (∼1550 cm^{-1}) and *III* (1400–1200 cm^{-1}) bands are also dependent on the secondary structural content; however, the structure–frequency correlation is much less straightforward. Consequently, these two bands are rarely used in structural studies (57).

Absorption of Amino Acid Side Chains. The absorption region of most side chains overlaps with the amide I spectral region (1610–1700 cm^{-1}), often causing difficulties in spectral interpretation (58). Only two amino acid side-chain groups are "interference-free", as they absorb in distinct spectral regions without overlapping with other groups. These are the protonated carboxyl groups of Asp and Glu, 1710–1790-cm^{-1} (deprotonated Asp and Glu side chains absorb strongly in the 1550–1580 cm^{-1} region, as well as near 1400 cm^{-1}), and the sulfhydryl group of Cys (2550–2600 cm^{-1}). Assignment of bands corresponding to other side chains is not easy and usually requires additional experiments. The most popular choice for these experiments is isotope labeling (change in the atomic mass results in a band shift according to Eq. 2-4-3). This result may be achieved by H/D exchange of labile hydrogen atoms, although in many cases this is not very helpful for a variety of reasons (e.g., too

many groups are affected by this exchange, resulting in too many changes in the spectrum). Alternatively, one can use uniform isotopic labeling of one amino acid, thus minimizing the number of band shifts.

Absorption of side chains is greatly influenced by their environment, which makes vibrational spectroscopy a sensitive tool to probe both intra- and intermolecular interactions. As such, vibrational spectroscopy often provides an opportunity to obtain information on protein properties that are difficult to discern with other biophysical methods. Only several of these features will be listed here (those that are important within the context of the discussion in the following chapters of the book). Since the bond vibration frequency depends on the electron density (in the form of k in Eq. 2-4-3), FTIR spectroscopy may be used to monitor *redox reactions* (59). Likewise, *metal chelation* by carboxylate groups, as well as the *protonation state* of amino acid side chains, often results in significant (and detectable) frequency shifts (57). Finally, bandwidth can be used as a semiquantitative measure of protein conformational heterogeneity (with flexible structures giving broader bands). Since the time scale of IR absorption measurements is very fast, it is being increasingly used for studies that require high temporal resolution (e.g., enzyme catalysis and charge transfer in redox-active systems) (60).

Fourier transform IR spectroscopy can also be used as a means to monitor protein HDX reactions. Replac-ing of ^1H in an N−H group with D (^2H) results in detectable frequency shifts (as expected from Eq. 2-4-3). Unfortunately, the measurements are complicated due to significant spectral interference (from side chains, as well as $^1H^2HO$), making such studies not as conclusive as those carried out with NMR or mass spectrometry (MS).

Finally, it should probably be mentioned again that water (being a strong IR absorber) often interferes with IR spectroscopic measurements. As a result, aqueous solutions of proteins and other macromolecules cannot easily be investigated using regular IR spectroscopic methods, necessitating the use of thin films and other solid samples.

Raman Spectroscopy. The technical difficulties mentioned above (interference of water absorption bands with those of proteins) can be largely overcome by using more sophisticated vibrational spectroscopic techniques (e.g., Raman spectroscopy). Although it provides information similar to that extracted from regular IR absorption spectra (i.e., normal modes of vibration), fundamental mechanistic differences (Raman is an inelastic scattering, NOT an absorption technique) lead to significant variation in the quality of information obtained from these experiments. While diminished H_2O and D_2O interference in Raman spectra is certainly a big advantage, Raman spectroscopy suffers several drawbacks that often limit its application. The S/N is typically inferior to that in the absorption IR spectra and quantitation is usually rather problematic (61). While most Raman studies target amide I and III bands (aimed at secondary and tertiary structure elucidation), there is growing interest in evaluating structural information that can be obtained from the side-chain bands (62).

UV Resonance Raman. If the excitation wavelength lies within a region of the protein electronic absorption, the spectrum will contain only the bands from the chromophore in resonance (off-resonance Raman scattering events would have a tremendously lower probability). Thus, UV resonance Raman (UVRR) spectroscopy provides a means to obtain a spectrum that contains signatures from certain amino acid residues (e.g., Trp and Tyr at 229-nm excitation) or protein backbone (amide $\pi \rightarrow \pi^*$ transition at 206 nm). The Raman spectra obtained using the 229-nm excitation are obviously dependent on the environment of Trp and Tyr residues within the protein. As such, they would reflect tertiary structural arrangement of the protein. The 206-nm excitation is also structurally informative, since the amide bond excitation gives rise to Raman spectra dominated by the amide vibrations whose characteristics (frequency, bandwidth, etc.) depend on protein secondary structure. Deconvolution of such spectra using a calibration set of "pure" secondary structural elements can provide quantitative information on the secondary structure content of proteins being examined (63). One may be compelled to draw parallels between far- and near-UV CD and UVRR utilizing 206 and 229-nm excitation, respectively.

Raman Optical Activity. Raman optical activity (ROA) spectroscopy can be viewed as a hybrid of CD and IR Raman spectroscopy. Current applications of ROA include secondary structure assignment (utilizing mostly amide I and III regions), evaluation of protein conformational heterogeneity (sometimes beyond that accessible by CD or NMR), as well as studies of (partially) disordered protein states (64).

2.4.3. Fluorescence: Monitoring Specific Dynamic Events

Fluorescence is another popular spectroscopic technique that is widely used to study macromolecular behavior in solution. The aromatic groups of three amino acids (tryptophan, tyrosine, and phenylalanine) offer intrinsic fluorescent probes of protein conformation, dynamics, and intermolecular interactions. Tryptophan is perhaps the most popular probe, since it occurs in one or a few positions in most proteins. The fluorescence of the indole chromophore is highly sensitive to the environment, making tryptophan an ideal choice for reporting protein conformational changes and interactions with other molecules (65). Tryptophan absorbs at 275–295 nm and emits at 320–350 nm (tyrosine

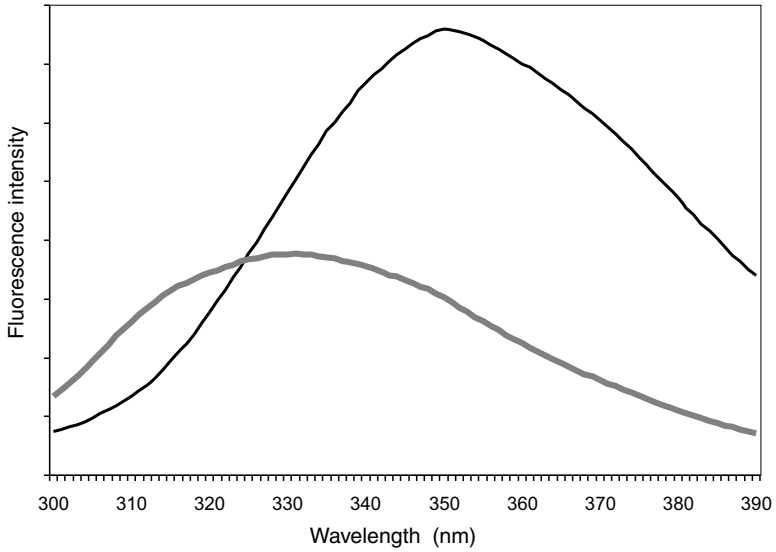


Figure 2.7. Tryptophan fluorescence emission spectrum of native and denatured states of cellular retinoic acid binding protein I (CRABP I), excited at 280 nm. The unfolded state (black line) shows a broad fluorescence signal centered ~350 nm, characteristic of solvent exposed tryptophan. By contrast the folded state (gray line) has a spectrum that is blue-shifted to 328 nm and significantly quenched (less intense) than that of the unfolded state. This is due to the specific quenching effect of tryptophan fluorescence by a proximal cysteine residue and is characteristic of native structure in this protein.

also fluoresces in this region, but its contribution is generally less significant). In principle, if the protein structure is known, then changes in tryptophan fluorescence could be interpreted in structural terms at atomic resolution (65). However, in practice the fluorescence signature of a protein arises from a large variety of contributions that cannot be adequately modeled.

In its simplest form, fluorescence can be used at equilibrium as a measure of structure due to the quenching effect of solvent exposure. In a completely denatured protein, tryptophan residues are exposed to surrounding solvent and fluorescence is quenched by collisions with solvent molecules. In contrast, a folded protein will generally have a tightly packed hydrophobic core in which aromatic residues are buried and sequestered from solvent (Fig. 2.7). This generally leads to an increased fluorescence signal from proteins in structured states, and generally a shift in the emission spectrum to higher frequency (blue shift), due to the reduced dielectric constant in the interior of a folded protein relative to solvent (66). Thus, measuring the fluorescence emission spectrum for a protein can be a useful measure of the extent of burial of aromatic residues. Since this will change during the folding of a protein, monitoring changes of fluorescence in a kinetic manner is a very sensitive probe of folding kinetics.

Another aspect of fluorescence that has been used extensively for study of protein structure and dynamics is the principle of fluorescence resonance energy transfer (FRET). Where there is overlap between the emission spectra of one fluorophore (donor) and the absorption spectrum of a second (acceptor), if the two moieties are sufficiently close in space there is the possibility of energy transfer between the two upon excitation of the donor. This manifests itself as a reduction in fluorescence of the donor and enhanced fluorescence of the acceptor, the efficiency of energy transfer depending on the distance between the donor and acceptor pair. Careful choice of chromophores, each pair of which have a different characteristic distance of 50% transfer efficiency, or Förster distance (67), allows these to be used as *molecular rulers* from which intermolecular distances (or simply distances between regions of a single macromolecule) can be measured. Dynamics in proteins can also be measured using FRET if the dynamic event involves changing the distance between the two fluorophores. An excellent source of information on the theory of the FRET technique may be found in (68).

Time-Resolved Fluorescence. Fluorescence emission is characterized by two parameters: the quantum yield* (Φ_0)

*Quantum yield is defined as a ratio of the number of emitted and absorbed photons and determines the efficiency of fluorescence for any given molecule.

and the lifetime of the excited state τ. Both parameters (Φ_0 and τ) are modulated by a number of intrinsic and extrinsic factors, most notably by the environment of the fluorophoric groups within the macromolecule. This effect is often used to evaluate macromolecular structure and dynamics in solution. Instead of re-emitting the photon, a number of other events can occur leading to the decay of the excited state (through a collision with another molecule or energy transfer to another group). The latter event is referred to as *quenching* and can be used to evaluate interaction of the fluorescent groups within the protein. The lifetime of the excited state of a single tryptophan varies from a few hundred picoseconds to 10 ns, and can be generally expressed as (65,69):

$$\frac{1}{\tau} = k_{\mathrm{r}} + k_{\mathrm{isc}} + k_{\mathrm{sol}} + \sum_i k_{\mathrm{q}}^i \qquad (2\text{-}4\text{-}5)$$

where k_{r} is the radiative rate constant, k_{isc} is the rate constant for the intersystem crossing, k_{sol} is the rate constant for solvent quenching, and k_{q}^i are rate constants for quenching through different mechanisms. Solvent quenching of the tryptophan fluorescence is typically in the range 10^7–$10^8\ \mathrm{s}^{-1}$ for unstructured polypeptides, but is significantly reduced if the protein conformation prevents solvent access to the indole group. Peptide bonds are another common quencher of tryptophan fluorescence. The rate constants of quenching by side chains exhibit significant variation depending on the mechanism of quenching and the particular side chain involved (69).

Since the tryptophan side chain is typically involved in multiple quenching interactions within the protein, the observed fluorescence decay kinetics are usually very complex even if only one tryptophan residue is present in the protein (69). In a pulsed fluorescence experiment, the emitted fluorescence is analyzed as a function of time following a short intense light pulse (excitation). Finite duration of the excitation pulse introduces another complication, as the measured total fluorescence intensity $I_{\mathrm{F}}(t)$ is actually a convolution of the excitation pulse profile $\mathrm{g}(t)$ and the intensity of radiation emitted by a single fluorescent particle $i_{\mathrm{F}}(t)$ (70):

$$I_{\mathrm{F}}(t) = \int_0^{\mathrm{t}} g(t') \cdot i_{\mathrm{F}}(t - t')dt' \qquad (2\text{-}4\text{-}6)$$

Deconvolution of $i_{\mathrm{F}}(t)$ based on the known shape of $\mathrm{g}(t)$ and measured $I_{\mathrm{F}}(t)$ can be carried out using an inverse FT procedure (71). If a fluorescent molecule is excited by polarized light, measurements of the polarization anisotropy of the emitted light can be used to study molecular tumbling (70,72).

Because of its superior sensitivity, fluorescence can be used to study structure and behavior of macromolecules present in solutions in minute concentrations.

Perhaps one of the most exciting recent developments in the field is the emergence of "single-molecule spectroscopy", a method of studying individual biopolymers under physiological conditions (73). Monitoring one molecule at a time often provides unique information on distribution functions of relevant observable properties, allowing subpopulations in a heterogeneous sample to be resolved (74–76).

2.5. OTHER BIOPHYSICAL METHODS TO STUDY MACROMOLECULAR INTERACTIONS AND DYNAMICS

2.5.1. Calorimetric Methods

Calorimetry is the best overall biophysical method to extract accurate thermodynamic data about the stability of proteins and binding interactions, and is the only method to measure the enthalpy change ΔH directly (77,78). When proteins fold and unfold, or when a protein binds a ligand, for instance, there will be a change in the free energy associated with the event. This temperature-dependent phenomenon can be described by

$$\Delta G(T) = \Delta H(T_{\mathrm{m}}) + \int_{T_{\mathrm{m}}}^{T} \Delta C_p dT - T\Delta S(T_{\mathrm{m}}) - T\int_{T_{\mathrm{m}}}^{T} \Delta C_p d\ln T \qquad (2\text{-}5\text{-}1)$$

Differential scanning calorimetry (DSC) measures the heat capacity change, ΔC_p, of a solution as a function of temperature (79–81). The DSC instruments contain two cells, to one of which is added a solution of the biomolecule(s) of interest and to the other (reference cell) an equal volume of solvent. The system is then heated adiabatically at a constant rate and the difference in power required to maintain the two cells at the same temperature represents the excess heat capacity of the sample. To obtain the temperature dependence of the heat capacity, the mass of analyte (m_a) and the partial specific volumes of analyte and solvent ($\bar{v}_a$ and $\bar{v}_s$) must be known. Then

$$C_p(T) = C_p^{\mathrm{solv}} \frac{\bar{v}_a}{\bar{v}_s} + \frac{\Delta C_p^{\mathrm{sol-solv}}}{m_a} \qquad (2\text{-}5\text{-}2)$$

where $\Delta C_p^{\mathrm{sol-solv}}$ is the measured heat capacity change in the calorimeter and C_p^{solv} is the heat capacity of the solvent alone.

When a protein is thermally denatured its heat capacity will generally change linearly with temperature until it begins to unfold, after which there is a large increase in $C_p(T)$ peaking at the midpoint of unfolding, T_{m}, as shown in

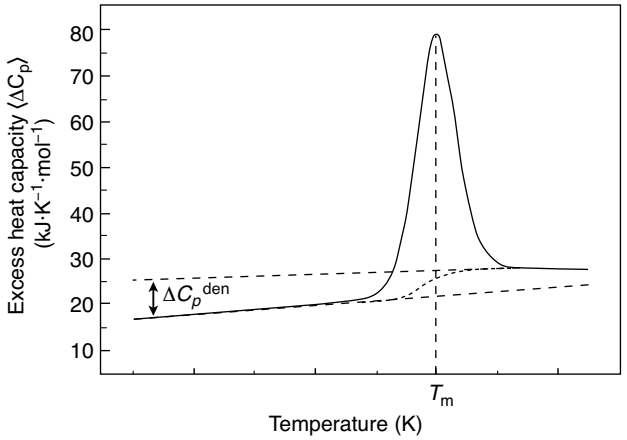


Figure 2.8. Calorimetric data obtained from DSC. The DSC thermogram showing the change in excess heat capacity with temperature as a protein undergoes thermal denaturation. Various fitting procedures allow parameters, such as the heat capacity change of denaturation (ΔC_p^{den}), and midpoint temperature of denaturation (T_{m}) to be extracted from the data. [Adapted with permission from (77). Copyright © 2001 John Wiley & Sons.]

Figure 2.8. The unfolded protein will have a higher heat capacity than the native protein that increases linearly with temperature. Deviations from the predicted C_p values for a known polypeptide chain may be indicative of residual structure in the thermally denatured state.

The calorimetric enthalpy and entropy changes for unfolding can be measured from integration under the curve for the DSC thermograms. Using the van't Hoff equation, an alternative enthalpy can be calculated assuming two-state behavior. If the ratio of the enthalpies calculated by both methods is unity, then this is indicative of a cooperative transition between folded and unfolded states. Deviation from unity may suggest either the presence of unfolding intermediates or else an irreversible unfolding process, depending on the direction of deviation.

Calorimetry can also be applied to study the thermodynamics of protein–ligand and protein–DNA association reactions (82,83). If the interaction is of sufficiently high affinity, then there will be a difference between the heat capacity of the complex versus the sum of those of the isolated components. By performing calorimetric measurements on the separate species and on the complex it is possible to obtain insight into thermodynamic forces that stabilize the interaction. However, care must be exercised since binding affinities are often very sensitive to external factors, such as pH and ionic strength (and pH is itself temperature dependent). Also, deconvolution of thermograms from protein complexes can be difficult, especially if each component of the system undergoes a thermal transition (84).

To investigate interactions between molecules, a more tractable approach is to use a different calorimetric method known as isothermal titration calorimetry (ITC). In this case, the system is held at constant temperature and one binding partner is titrated into a solution containing the second binding partner in a highly controlled manner. As each aliquot of titrant is added, the heat change is measured relative to a reference cell by determining the amount of power required to keep a constant temperature difference between the two cells.

Figure 2.9*a* shows a typical titration profile. Initially, there is a large heat change as titrant is added, but as binding approaches saturation the heat change diminishes with further addition until finally there is no further change. From these experiments, binding curves can be constructed to obtain association constants. In addition, the total heat absorbed or released by the system is the enthalpy change

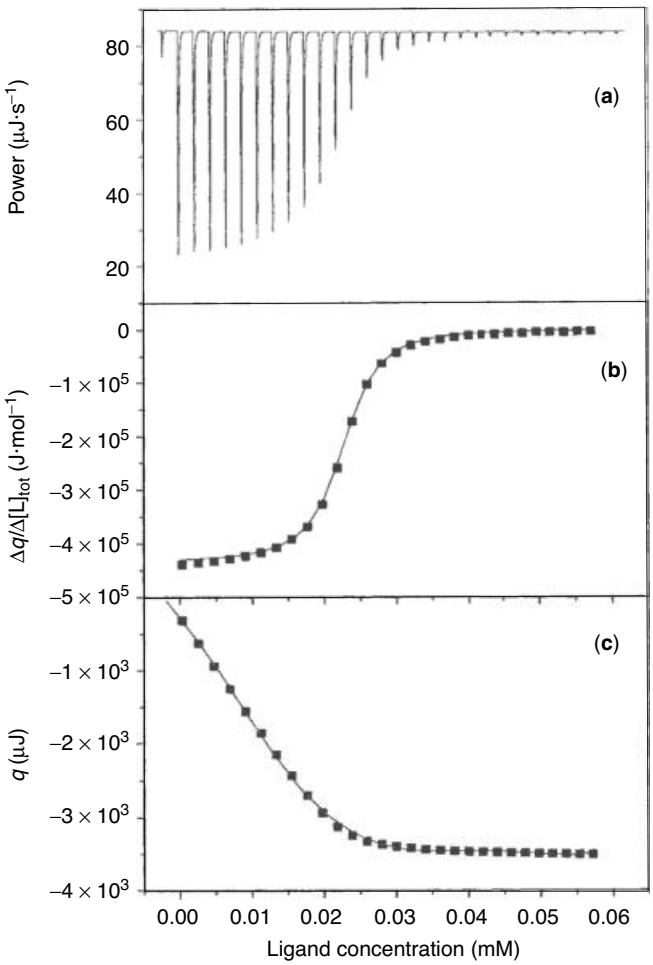


Figure 2.9. (*a*) Differential power signal upon titrating complementary strands of 16 bp DNA at 30°C measured by ITC. Upon each addition, heat is absorbed leading to a spike in the power trace. Integration yields titration curves (*b,c*) of heat change (Δq) versus ligand concentration ([L]), from which binding constants and enthalpies of association are determined. [Adapted with permission from (77). Copyright © 2001 John Wiley & Sons.]

upon binding. Once again this calorimetric value can be compared with van't Hoff enthalpies obtained by other methods to determine whether the association is a cooperative process. By performing ITC titrations as a function of temperature, the change in heat capacity of binding can also be determined. This represents a measure of the burial of solvent-exposed surface area occurring when the two interaction partners bind. Again this can also be calculated empirically, and deviations between theoretical and experimental values can provide useful information about the nature of the binding interactions. Only in the case of a simple lock-and-key binding mechanism would one expect the two values to coincide. With an induced fit mechanism, the rearrangements required to bind ligand would be associated with an additional change in ΔC_p. This effect is likely even further accentuated in the case of intrinsically disordered proteins, where in order to bind the ligand effectively the protein folds around the template ligand, thus producing a large heat change.

2.5.2. Analytical Ultracentrifugation

The technique of analytical ultracentrifugation has been used for many years to study the molecular weights, hydrodynamic properties, and solution interactions in macromolecules (81,85–87). The method takes advantage of the effect of diffusion of molecules within a solution that can be balanced by the tendency of the molecules to sediment under the influence of a centrifugal force. While NMR and crystallography can yield high-resolution structural data, the ultracentrifuge is extremely versatile and can provide details on size and shape of the sedimenting species. It can also be used to obtain equilibrium association constants and to measure the extent of aggregation states and conformational changes.

In a spinning rotor, a particle with mass m experiences a sedimenting force F_s that depends on the angular velocity (ω) of the rotor, and the radial distance (r) from the axis of rotation:

$$F_s = m\omega^2 r \tag{2-5-3}$$

This force is balanced by a buoyant force F_b that is a function of the solvent density (ρ) and the partial specific volume of the solute ($\bar{v}$), and a frictional force F_f that depends on the velocity of the particle (u) and its shape and size, denoted by the frictional coefficient, f:

$$F_b = -m\bar{v}\rho\omega^2 r \tag{2-5-4}$$

$$F_f = -fu \tag{2-5-5}$$

These forces balance and, as such, these equations may be combined to produce the Svedberg equation:

$$\frac{MD(1-\bar{v}\rho)}{RT} = \frac{u}{\omega^2 r} \equiv s \tag{2-5-6}$$

where the diffusion coefficient, $D = RT/N_A f$, N_A is Avogadro's number, M is the molecular weight of the particle, and s is the sedimentation coefficient, which has the unit of svedbergs ($S = 10^{-13}$ s).

The experimental setup of the analytical ultracentrifuge involves a rotor that can be rotated at a very precisely controlled speed and temperature. In order to sediment macromolecules in a reasonable time frame the centrifuge is normally capable of speeds up to 60,000 rpm and the rotor chamber is evacuated to reduce friction and turbulence. Additionally, and most importantly, an optical system is present that can precisely measure the radial concentration of analyte across the observation cell. Commercially available instrumentation normally measures optical absorbance or interference due to the difference in refractive index between sample and solute, although fluorometric detection systems are also available (88).

The sample is introduced into one sector of a two-sector cell while the adjacent sector contains only the reference solvent, precisely matched in concentration and ionic strength to that containing the analyte. The cell is then aligned radially in the centrifuge such that sedimentation under centrifugal force can be measured by the variation of solute concentration as a function of radial distance from the rotor axis. The optical system is precisely synchronized to the rotation of the rotor so that observation occurs at the same position in the cell with each pass of the rotor through the light path. In older systems, absorbance or interference was measured using film and then measuring the intensity of the photographic image or counting interference fringes manually, but with the advent of CCD cameras and computer imaging systems, this is all now computer controlled.

Two types of experiment can be performed in an analytical ultracentrifuge, the first of which is known as sedimentation velocity. This method can be used to rapidly assess the number of species present in a mixture, whether there are interacting species and the sedimentation and diffusion coefficients. It is often used as a precursor to more detailed study by sedimentation equilibrium (see below).

An initially uniform solution of analyte is subjected to a sufficiently high angular velocity that the solute begins to sediment rapidly toward the bottom of the cell (Fig. 2.10*a*). As solute is depleted from the meniscus region of the sample column, a sharp boundary forms that shifts toward the bottom of the cell with time, and also broadens due to diffusion. The rate of sedimentation depends on the molecular weight and shape of the solute molecules. From these experiments, the rate of broadening of the boundary can be used to calculate the diffusion coefficient, and the rate of sedimentation gives the sedimentation coefficient, from

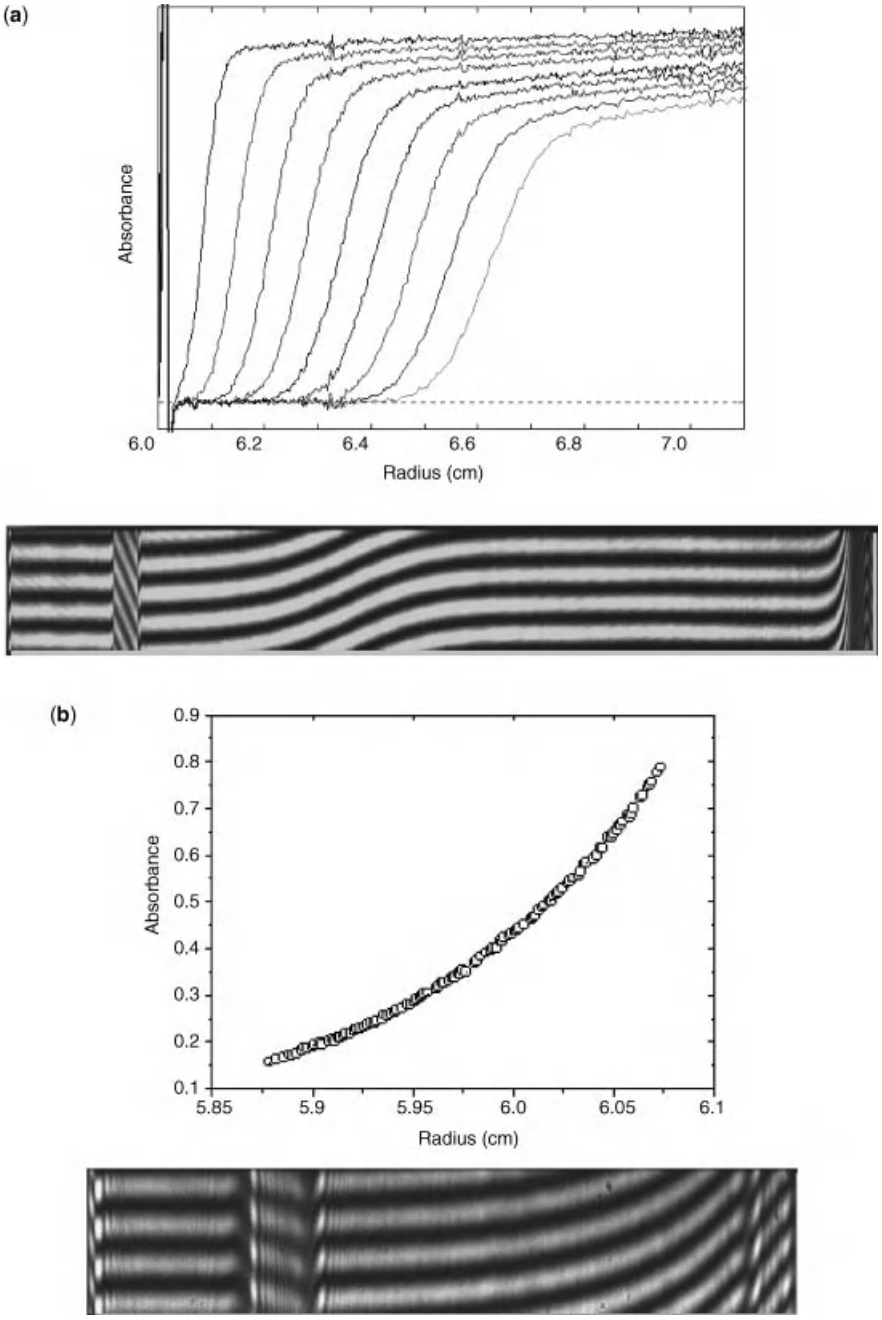


Figure 2.10. Typical traces obtained from (*a*) sedimentation velocity, and (*b*) sedimentation equilibrium experiments using analytical ultracentrifugation. In (*a*) the change in absorbance along the length of the observation cell is plotted, each curve representing a different time point. As the experiment proceeds, absorbing material begins to sediment, leading to a depletion in the absorbance in the bulk solution and the sedimenting boundary shifts along the cell. In (*b*), equilibrium has been reached where sedimentation and diffusion forces are exactly balanced, leading to a distribution of material throughout the cell whose properties can be used to determine molecular parameters, such as molecular weight and dissociation constants.

which the molecular weight of the analyte can be calculated. The latter can be determined according to the equation:

$$s = \frac{dr_{\text{bnd}}/dt}{\omega^2 r} \tag{2-5-7}$$

$$\ln(r_{\text{bnd}}/r_{\text{m}}) = s\omega^2 t \tag{2-5-8}$$

where r_{bnd} is the radial distance of the boundary from the axis, and r_{m} is the distance of the solvent meniscus. Thus, the sedimentation coefficient can be determined from a plot of $\ln(r_{\text{bnd}})$ versus time. In practice, a variety of alternative methods have been developed to determine sedimentation coefficients that account for such factors as nonideality of the solution, radial dilution, self-association, concentration dependence of the sedimentation coefficient, and even sedimentation occurring during the initial rotor acceleration. Perhaps one of the most elegant of these involves computational numerical solution of the Lamm equation (86,89,90):

$$\frac{dc(r,t)}{dt} = \frac{1}{r}\frac{d}{dr}\left[rD\frac{dc(r,t)}{dr} - s\omega^2 r^2 c(r,t)\right] \tag{2-5-9}$$

which represents a general equation for sedimentation of a particle of sedimentation coefficient s and diffusion coefficient D. Using the wealth of time-dependent concentration data obtained from the ultracentrifuge, Eq. 2-5-9 can be rapidly solved computationally to fit the data and obtain values for the coefficients of the sedimenting species. This method has been successfully used to measure molecular weight distributions ranging from small molecules all the way up to viral capsids (>1 MDa).

In the case of multiple populations present in the sample, the individual components will sediment at different rates according to their molecular weight and shape. If the components are sufficiently different, multiple boundaries can be observed, enabling extraction of parameters for the individual species. Note that the sedimentation coefficients are not additive. For instance, ribosomes consist of two subunits that have molecular weights corresponding to 40S and 60S. However the intact ribosome has a sedimentation coefficient of 80S.

The second type of experiment involves much lower rotor speeds, and instead of simply sedimenting the solute molecules, they are allowed to come to equilibrium where sedimentation forces are exactly balanced by diffusion. In this case, the boundaries between solvent and solute reach an equilibrium radial distribution, solute concentration increases exponentially with increasing radius, and becomes invariant with time (Fig. 2.10*b*). Because it is an equilibrium method, this technique can be used to obtain thermodynamic information about interacting species. In the case of a reversible association, such as protein–protein or protein–nucleotide binding, one can obtain equilibrium dissociation constants. It can also be used to detect the existence of nonspecific associations and/or aggregation events. The disadvantage is that, whereas a sedimentation velocity experiment can be completed in a matter of hours, it can sometimes take several days for the system to reach a satisfactory equilibrium in the ultracentrifuge. In many cases, this can be circumvented by using short solution columns, but at the potential expense of resolution from using only a small portion of the observation cell for measurements.

At equilibrium, the buoyant molecular weight of a solvent may be derived from Eq. 2-5-6:

$$M = \frac{2RT}{(1-\bar{v}\rho)\omega^2}\frac{d(\ln c)}{dr^2} \tag{2-5-10}$$

Thus, at equilibrium the molecular weight of a single species can be determined from the slope of a plot of $\ln c$ versus r^2. In practice, however, very few biomolecules exist purely as monomers and the real power of equilibrium sedimentation lies in its ability to probe these intermolecular interactions. As the concentration varies across the radial length of the cell, so too will the degree of association by the law of mass action. At the top of the cell, the concentration of oligomer is negligible so monomer molecular weights can be extracted, whereas toward the bottom of the cell the increased molecular weight resulting from associations can be probed.

Normally, a complete data set will include experiments at a series of concentrations, rotor speeds, and even temperatures, from which a global fitting procedure can be applied to model different interaction scenarios (monomer–dimer, monomer–tetramer, etc.). Other treatments can determine whether these associations are truly reversible or are, in fact, aggregation events, and also electrostatic effects that can lead to solution nonideality, and hence ambiguity of the results. These complicated procedures are beyond the scope of this text, but excellent review articles and introductory texts are available (85,86,90–93). Careful and thorough analysis, however, can be used to gain extremely valuable details about molecular weights, thermodynamics of association, and has even been applied to investigate the effects of detergents on membrane proteins and the effects of macromolecular crowding as is likely the case in the intracellular environment.

2.5.3. Surface Plasmon Resonance

This method has become very popular for the study of protein–ligand and protein–protein interactions. The technique relies on a phenomenon when incident light passes between a high refractive index (n_1) and low refractive index (n_2, where $n_2 < n_1$) medium when there is an electrically conducting surface at the interface (94–96). Above a critical

angle (θ) of incident light, the beam is reflected at the interface back into the high refractive index medium. However, in doing so, it also sets up an electric field in the conducting surface that can lead to a conversion of photon energy into plasmon energy at the surface in a manner that is critically dependent on the refractive index (n_2) at the surface. If the wave vectors of the photon (k_x) and plasmon (k_{sp}) energies are equal in magnitude and direction, then a resonant effect occurs (known as surface plasmon resonance, SPR) between the two that manifests itself as a reduction in intensity of the reflected light:

$$k_x = \frac{2\pi}{\lambda} n_1 \sin\theta \qquad (2\text{-}5\text{-}11)$$

$$k_{sp} = \frac{2\pi}{\lambda} \sqrt{\frac{n_{\text{gold}}^2 n_2^2}{n_{\text{gold}}^2 + n_2^2}} \qquad (2\text{-}5\text{-}12)$$

where n_{gold} is a constant, the refractive index of the gold layer, λ is the wavelength of incident light. The parameter k_x can be tuned to match k_{sp} either by varying the wavelength or angle of the incident beam. In practice, the angle θ is varied and the angle of minimum reflectance intensity due to the SPR effect is measured. The refractive index (n_2) at the surface can then be calculated by equating 2-5-11 and 2-5-12 ($k_{sp} = k_x$). If a protein is immobilized very close to the surface, then binding of another protein or ligand will cause a change in concentration, and hence refractive index (n_2) at the surface. Thus, the SPR phenomenon can be used as a very sensitive probe of concentration changes, and hence protein–ligand interactions.

A glass slide coated with a thin layer of gold can be decorated at its surface with the biomolecule of interest. A variety of chemistries have been developed to allow this to happen readily, for instance, a thin layer of carboxymethylated dextran allows attachment of proteins via a variety of reactive side chains (amine, thiol, carboxylate), while also providing a hydrophilic environment to accommodate the binding interaction. More specific coatings, such as streptavidin, can be used to select biotinylated proteins or nickel nitrilotriacetate will specifically bind proteins with histidine tags.

The plate is then placed into the instrument with the protein–decorated surface exposed to a liquid flow chamber. As buffer containing the interaction partner of the immobilized protein is flowed over the surface, protein–protein interactions will occur, and the binding event causes a change in the refractive index of the surface that is dependent on the concentration of bound species. This refractive index change is detected as a change in the SPR response, and can thus be used quantitatively to measure the binding constant. This technique is extremely sensitive and versatile and can be used to screen interactions of proteins with ligands. Since the target protein is linked either covalently or by a high-affinity interaction with the surface, ligands can in general be washed off by adjusting buffer conditions, thus regenerating the surface so that the experiment can be repeated with different concentrations of ligand or else completely different molecules. Some information can be obtained about the kinetics of binding from the time course for development of the SPR response, and thermodynamic binding data can be extracted from the concentrations of bound ligand at steady state. However, care must be taken to ensure that the bound proteins are in the correct orientation on the surface to enable efficient binding of ligand, and also that the immobilization process does not in some way compromise the structural integrity or ligand-binding efficacy of the target protein.

2.5.4. Size Exclusion Chromatography (Gel Filtration)

The means by which a macromolecule passes through a chromatography column packed with silica gel depends empirically on its size and shape. Larger molecules will be excluded from the pores formed within the gel bed, and hence will pass through the column rapidly. On the other hand, small molecules can enter at least partially into these cavities, and hence passage of these molecules will be retarded. Size exclusion chromatography, SEC (also known as gel filtration or gel permeation chromatography) is a powerful technique for determining molecular weight and distributions in random coil synthetic polymers. However, it is not trivial to characterize the molecular weight of folded protein molecules since shape also has a significant effect. Regions of a protein with an extended conformation may become entangled in the gel pores, and hence cannot pass through the column as rapidly as one would predict from their molecular weight. Similarly, molecules with unusual charge distributions or patches of hydrophobic residues may interact with the column resin in unpredictable ways.

Nevertheless, gel filtration has uses for studying protein–protein interactions. Clearly, multimeric forms should have a different mobility than the monomer. With careful calibration and taking into account the many caveats, gel filtration can be a helpful empirical tool for the investigation of protein size and intermolecular interactions. In Chapter 9, we will see an example of how SEC is used in combination with MS to study protein aggregation processes. In this case, protein oligomers are separated from each other based on their size (i.e., number of monomer units). Size-based separation can be also applied, at least in principle, to separating folded proteins from their less structured forms. Figure 2.11 shows SEC profiles of two forms of a small glycoprotein interferon β1a, intact and alkylated. As we will see in Chapters 4 and 5, alkylation of the single cysteine residue in this protein results in partial unfolding, and SEC clearly shows shorter retention time, consistent with the (partial) loss of compactness in solution.

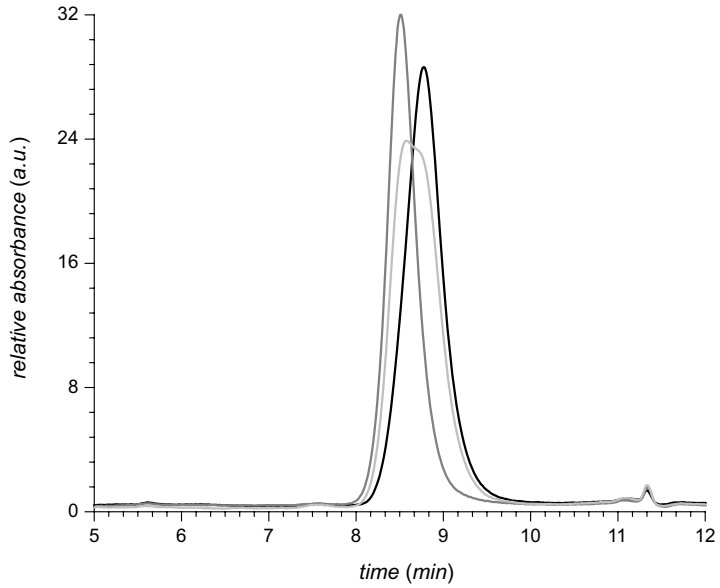


Figure 2.11. Size exclusion chromatography elution profiles of intact (black trace) and NEM-alkylated form (gray trace) of interferon β1a. Light gray trace corresponds to the incompletely alkylated protein. [Adapted with permission from (97). Copyright © 2008 American Chemical Society.]

One thing that needs to be kept in mind when interpreting the results of SEC measurements is that fast interconversion between different forms of the protein (e.g., unfolding–refolding or association–dissociation) during the chromatographic run may influence the outcome of such an experiment by generating a single peak, whose retention time will be indicative of how much time the protein molecule spends in each specific conformation during its passage through the column). This phenomenon may be especially detrimental in the analysis of reversible protein aggregates and, more generally, multiunit associations, since SEC measurements inevitably involve sample dilution, which shifts the equilibrium toward dissociation (98).

2.5.5. Electrophoresis

Electrophoresis involves applying an electrical potential across a thin (usually) polyacrylamide gel held between glass plates, to which samples of analyte biomolecules have been applied at one end. Under the influence of the electric field, the analyte will enter the gel and pass through it depending on a combination of molecular weight, molecular shape, and charge on the molecule. The most popular method involves denaturing the protein sample by boiling with β-mercaptoethanol and the anionic detergent sodium dodecyl sulfate (SDS). The former reagent reduces all disulfide bridges, while the latter causes complete denaturation and encapsulates the proteins in detergent micelles that have an approximately constant charge-to-mass ratio. Protein molecules will thus migrate through the gel matrix in a predictable manner, with smaller molecules passing through the gel more rapidly and high molecular weight ones more slowly. Once the samples have been separated, but before they pass completely through the gel, the electric field is removed and the proteins may be visualized by staining with a protein binding dye. By comparison with calibration standards, the molecular weight can be determined with some degree of precision.

Since this electrophoretic technique employs denaturing conditions, it is not useful for looking at biologically relevant macromolecular interactions. It is, however, possible to use nondenaturing conditions and apply native proteins to the gel. Due to the above dependences on size, shape, and charge, interpretation of the mobility of analyte through a native gel is not trivial. Normally, it is necessary to run a suite of gels with different degrees of cross-linking in the gel matrix to obtain useful data. Once again, this is a simple empirical technique that can be useful for initial screening of proteins and their interactions, but requires very careful calibration to obtain meaningful information.

Another electrophoretic technique that has been used to evaluate protein conformations and transitions among them is capillary electrophoresis (CE). As is the case with gel electrophoresis discussed earlier, the measurements are based on separating proteins according to their mobilities, which are determined by both total charge and the size of the protein molecule. Since the latter can be correlated with protein conformation (the degree of compactness), protein mobility data derived from CE measurements can be

indicative of protein conformation. Unlike gel electrophoresis, CE measurements are carried out in solution, thus it is much easier to perform experiments under physiologically relevant conditions (99). A significant advantage of CE is that it requires very small amounts of the protein sample. This technique has been shown to be able to resolve conformational intermediates, including protein oligomers (100). An interested reader is referred to an old, but nonetheless very useful review on this subject (99).

2.5.6. Affinity Chromatography

Affinity chromatography (AC) is another separation technique that has gained prominence in protein extraction/purification, but is also used in biophysical experiments (101). The separation is based upon the interaction of molecules immobilized to a surface (ligands) with their partners dissolved in the mobile phase (analyte). The interactions are usually highly specific (hence the use of AC in protein extraction from complex matrices), but the interaction strength can be modulated by a variety of extrinsic factors. Biophysical applications of AC may include discovery of novel binding partners for a given ligand and ranking of affinities within a set of target molecules toward a particular ligand, among others. A very detailed discussion of the technical aspects of affinity chromatography can be found in a recent review article by Urh et al. (101).

This chapter briefly reviewed a number of the biophysical techniques routinely used in the laboratory to study protein size, shape, structure, and dynamics. Some of these are very low resolution, others are extremely high resolution, providing anything from a quick and general view of protein associations right through to a picture at the atomic level. In the following chapters, we will demonstrate some of the many applications of MS to answer biological questions. In many ways, these techniques are complementary to the methods described herein, providing confirmatory evidence and in many cases unique information. No single biophysical method is sufficient to fully describe a system, but in Chapters 3–9 we will see the power of MS, both now and in the future, in the biophysical arena(101).

REFERENCES

1. Rhodes, G. (2000) *Crystallography made crystal clear. A guide for users of macromolecular models.*, Academic Press, Inc., San Diego.
2. Schmidt, A. and Lamzin, V. S. (2002) Veni, vidi, vici - atomic resolution unravelling the mysteries of protein function, *Curr. Opin. Struct. Biol. 12*, 698–703.
3. Longhi, S., Czjzek, M., and Cambillau, C. (1998) Messages from ultrahigh resolution crystal structures, *Curr. Opin. Struct. Biol. 8*, 730–737.
4. Lario, P. I. and Vrielink, A. (2003) Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution, *J. Am. Chem. Soc. 125*, 12787–12794.
5. Arora, A. and Tamm, L. K. (2001) Biophysical approaches to membrane protein structure determination, *Curr. Opin. Struct. Biol. 11*, 540–547.
6. Werten, P. J., Remigy, H. W., de Groot, B. L., Fotiadis, D., Philippsen, A., Stahlberg, H., Grubmuller, H., and Engel, A. (2002) Progress in the analysis of membrane protein structure and function, *FEBS Lett. 529*, 65–72.
7. Stahlberg, H., Fotiadis, D., Scheuring, S., Remigy, H., Braun, T., Mitsuoka, K., Fujiyoshi, Y., and Engel, A. (2001) Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins, *FEBS Lett. 504*, 166–172.
8. Read, R. J. (1996) As MAD as can be, *Structure 4*, 11–14.
9. Mozzarelli, A. and Rossi, G. L. (1996) Protein function in the crystal, *Annu. Rev. Biophys. Biomol. Struct. 25*, 343–365.
10. Moffat, K. (2001) Time-resolved biochemical crystallography: a mechanistic perspective, *Chem. Rev. 101*, 1569–1581.
11. Petsko, G. A. and Ringe, D. (2000) Observation of unstable species in enzyme-catalyzed transformations using protein crystallography, *Curr. Opin. Chem. Biol. 4*, 89–94.
12. Murphy, R. M. (1997) Static and dynamic light scattering of biological macromolecules: what can we learn?, *Curr. Opin. Biotechnol. 8*, 25–30.
13. Yamaguchi, T. and Adachi, K. (2002) Hemoglobin equilibrium analysis by the multiangle laser light- scattering method, *Biochem. Biophys. Res. Commun. 290*, 1382–1387.
14. Murphy, R. M. and Pallitto, M. M. (2000) Probing the kinetics of beta-amyloid self-association, *J. Struct. Biol. 130*, 109–122.
15. Wall, M. E., Gallagher, S. C., and Trewhella, J. (2000) Large-scale shape changes in proteins and macromolecular complexes, *Annu. Rev. Phys. Chem. 51*, 355–380.
16. Doniach, S. (2001) Changes in biomolecular conformation seen by small angle X-ray scattering, *Chem. Rev. 101*, 1763–1778.
17. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys. J. 76*, 2879–2886.
18. Chen, L., Hodgson, K. O., and Doniach, S. (1996) A lysozyme folding intermediate revealed by solution X-ray scattering, *J. Mol. Biol. 261*, 658–671.
19. Segel, D. J., Eliezer, D., Uversky, V., Fink, A. L., Hodgson, K. O., and Doniach, S. (1999) Transient dimer in the refolding kinetics of cytochrome c characterized by small-angle X-ray scattering, *Biochemistry 38*, 15352–15359.
20. Unger, V. M. (2001) Electron cryomicroscopy methods, *Curr. Opin. Struct. Biol. 11*, 548–554.
21. Chiu, W., McGough, A., Sherman, M. B., and Schmid, M. F. (1999) High-resolution electron cryomicroscopy of macromolecular assemblies, *Trends Cell. Biol. 9*, 154–159.
22. Amos, L. A. (2000) Focusing-in on microtubules, *Curr. Opin. Struct. Biol. 10*, 236–241.

23. Frank, J. (2002) Single-particle imaging of macromolecules by cryo-electron microscopy, *Annu. Rev. Biophys. Biomol. Struct. 31*, 303–319.
24. Saibil, H. R. (2000) Conformational changes studied by cryo-electron microscopy, *Nat. Struct. Biol. 7*, 711–714.
25. Steven, A. C. and Aebi, U. (2003) The next ice age: cryo-electron tomography of intact cells, *Trends Cell Biol. 13*, 107–110.
26. Marsh, B. J., Mastronarde, D. N., Buttle, K. F., Howell, K. E., and McIntosh, J. R. (2001) Inaugural Article: Organellar relationships in the Golgi region of the pancreatic beta cell line, HIT-T15, visualized by high resolution electron tomography, *Proc. Natl. Acad. Sci. USA 98*, 2399–2406.
27. Baumeister, W. (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm, *Curr. Opin. Struct. Biol. 12*, 679–684.
28. Stoffler, D., Feja, B., Fahrenkrog, B., Walz, J., Typke, D., and Aebi, U. (2003) Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport, *J. Mol. Biol. 328*, 119–130.
29. Grunewald, K., Medalia, O., Gross, A., Steven, A. C., and Baumeister, W. (2003) Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding, *Biophys. Chem. 100*, 577–591.
30. Medalia, O., Weber, I., Frangakis, A. S., Nicastro, D., Gerisch, G., and Baumeister, W. (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography, *Science 298*, 1209–1213.
31. Byron, O. and Gilbert, R. J. (2000) Neutron scattering: good news for biotechnology, *Curr. Opin. Biotechnol. 11*, 72–80.
32. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc. Natl. Acad. Sci. USA 94*, 12366–12371.
33. Losonczi, J. A. and Prestegard, J. H. (1998) Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules, *J. Biomol. NMR 12*, 447–451.
34. Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium (vol 278, pg 1111, 1997), *Science 278*, 1697–1697.
35. Woodward, C. (1993) Is the slow-exchange core the protein-folding core, *Trends Biochem. Sci. 18*, 359–360.
36. Li, R. H. and Woodward, C. (1999) The hydrogen exchange core and protein folding, *Protein Sci. 8*, 1571–1590.
37. Parker, M. J. and Marqusee, S. (2001) A kinetic folding intermediate probed by native state hydrogen exchange, *J. Mol. Biol. 305*, 593–602.
38. Englander, S. W. (2000) Protein folding intermediates and pathways studied by hydrogen exchange, *Annu. Rev. Biophys. Biomol. Struct. 29*, 213–238.
39. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) Hydrogen exchange: The modern legacy of Linderstrøm-Lang, *Protein Sci. 6*, 1101–1109.
40. Raschke, T. M. and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions, *Nat. Struct. Biol. 4*, 298–304.
41. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH, *Nat. Struct. Biol. 3*, 782–787.
42. Bai, Y. W., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein-folding intermediates - native-state hydrogen-exchange, *Science 269*, 192–197.
43. Roder, H., Elove, G. A., and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome-c by H-exchange labeling and proton NMR, *Nature (London) 335*, 700–704.
44. Miranker, A., Radford, S. E., Karplus, M., and Dobson, C. M. (1991) Demonstration by NMR of folding domains in lysozyme, *Nature (London) 349*, 633–636.
45. Miranker, A., Robinson, C. V., Radford, S. E., and Dobson, C. M. (1996) Investigation of protein folding by mass spectrometry, *FASEB J. 10*, 93–101.
46. Balbach, J., Forge, V., Lau, W. S., vanNuland, N. A. J., Brew, K., and Dobson, C. M. (1996) Protein folding monitored at individual residues during a two-dimensional NMR experiment, *Science 274*, 1161–1163.
47. Balbach, J., Forge, V., Vannuland, N. A. J., Winder, S. L., Hore, P. J., and Dobson, C. M. (1995) Following protein-folding in real-time using NMR-spectroscopy, *Nat. Struct. Biol. 2*, 865–870.
48. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1995) *Protein NMR spectroscopy: principles and practice*, Academic Press, San Diego.
49. Levitt, M. H. (2001) *Spin dynamics: basics of nuclear magnetic resonance*, John Wiley & Sons, Inc., Chichester; New York.
50. Sreerama, N. and Woody, R. W. (2000) Circular dichroism of peptides and proteins, In *Circular Dichroism: principles and applications.* (Berova, N., Nakanishi, K., and Woody, R. W. Eds.) 2nd ed., pp 601–620, Wiley–VCH, New York.
51. Sreerama, N., and Woody, R. W. (2003) Structural composition of beta-I- and beta-II-proteins, *Protein Sci. 12*, 384–388.
52. Sreerama, N., Venyaminov, S., and Woody, R. (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy, *Protein Sci. 8*, 370–380.
53. Pelton, J. T. and McLean, L. R. (2000) Spectroscopic methods for analysis of protein secondary structure., *Anal. Biochem. 277*, 167–176.
54. Sreerama, N., Manning, M. C., Powers, M. E., Zhang, J. X., Goldenberg, D. P., and Woody, R. W. (1999) Tyrosine, phenylalanine, and disulfide contributions to the circular

dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor, *Biochemistry 38*, 10814–10822.

55. Johnson, W. C. (2000) CD of nucleic acids, In *Circular dichroism: principles and applications* (Berova, N., Nakanishi, K., and Woody, R. W., Eds.) 2nd ed., pp 703–718, Wiley–VCH, New York.
56. Maurizot, J. C. (2000) Circular dichroism of nucleic acids: nonclassical conformations and modified oligonucleotides, In *Circular dichroism: principles and applications* (Berova, N., Nakanishi, K., and Woody, R. W., Eds.) 2nd ed., pp 719–740, Wiley–VCH, New York.
57. Barth, A. and Zscherp, C. (2002) What vibrations tell us about proteins, *Q. Rev. Biophys. 35*, 369–430.
58. Barth, A. (2000) The infrared absorption of amino acid side chains, *Prog. Biophys. Mol. Biol. 74*, 141–173.
59. Breton, J. (2001) Fourier transform infrared spectroscopy of primary electron donors in type I photosynthetic reaction centers, *Biochim. Biophys. Acta 1507*, 180–193.
60. Zscherp, C. and Barth, A. (2001) Reaction-induced infrared difference spectroscopy for the study of protein reaction mechanisms, *Biochemistry 40*, 1875–1883.
61. Thomas, G. J., Jr. (1999) Raman spectroscopy of protein and nucleic acid assemblies, *Annu. Rev. Biophys. Biomol. Struct. 28*, 1–27.
62. Thomas, G. J., Jr. (2002) New structural insights from Raman spectroscopy of proteins and their assemblies, *Biopolymers 67*, 214–225.
63. Chi, Z., Chen, X. G., Holtz, J. S., and Asher, S. A. (1998) UV resonance Raman-selective amide vibrational enhancement: quantitative methodology for determining protein secondary structure, *Biochemistry 37*, 2854–2864.
64. Barron, L. D., Hecht, L., Blanch, E. W., and Bell, A. F. (2000) Solution structure and dynamics of biomolecules from Raman optical activity, *Prog. Biophys. Mol. Biol. 73*, 1–49.
65. Chen, Y. and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins, *Biochemistry 37*, 9976–9982.
66. Lakowicz, J. R. (1999) *Principles of fluorescence spectroscopy*, 2nd ed., Kluwer Academic/Plenum, New York.
67. Förster, T. (1948) Intermolecular energy migration and fluorescence, *Ann. Phys. (Leipzig) 2*, 55–75.
68. van der Meer, B. W., Coker, G., III, and Chen, S.-Y. (1994) *Resonance energy transfer: theory and data*, VCH, New York.
69. Engelborghs, Y. (2001) The analysis of time resolved protein fluorescence in multi-tryptophan proteins, *Spectrochim. Acta A: Mol. Biomol. Spectrosc. 57*, 2255–2270.
70. Daune, M. (1999) *Molecular biophysics: structures in motion*, Oxford University Press, Oxford; New York.
71. Kauppinen, J. and Partanen, J. (2001) *Fourier transforms in spectroscopy*, 1st ed., Wiley–VCH, Berlin; New York.
72. Fersht, A. (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, W.H. Freeman, New York.
73. Kelley, A. M., Michalet, X., and Weiss, S. (2001) Chemical physics: single-molecule spectroscopy comes of age, *Science 292*, 1671–1672.
74. Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy, *Nat. Struct. Biol. 7*, 724–729.
75. Kohler, J. (2001) Optical spectroscopy of individual objects, *Naturwissenschaften 88*, 514–521.
76. Michalet, X., Kapanidis, A. N., Laurence, T., Pinaud, F., Doose, S., Pflughoefft, M., and Weiss, S. (2003) The power and prospects of fluorescence microscopies and spectroscopies, *Annu. Rev. Biophys. Biomol. Struct. 32*, 161–182.
77. Jelesarov, I. and Bosshard, H. R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition., *J. Mol. Recognit. 12*, 3–18.
78. Ladbury, J. E. and Chowdry, B. Z., (Eds.) (1998) *Biocalorimetry: applications of calorimetry in the biological sciences.*, John Wiley & Sons, Inc., New York.
79. Freire, E. (1995) Differential scanning calorimetry, *Methods Mol. Biol. 40*, 191–218.
80. Privalov, P. L. and Potehkin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins., *Methods Enzymol. 131*, 4–51.
81. Sturtevant, J. M. (1987) Biochemical applications of differential scanning calorimetry., *Proc. Natl. Acad. Sci. USA 74*, 2236–2240.
82. Brandts, J. F. and Lin, L. N. (1990) Study of strong to ultratight protein interactions using differential scanning calorimetry, *Biochemistry 29*, 6927–6940.
83. Breslauer, K. J., Freire, E., and Straume, M. (1992) Calorimetry: a tool for DNA and ligand-DNA studies, *Methods Enzymol. 211*, 533–567.
84. Freire, E. (1994) Statistical thermodynamic analysis of differential scanning calorimetry data: structural deconvolution of heat capacity function of proteins, *Methods Enzymol. 240*, 502–530.
85. Laue, T. M. and Stafford, W. F., 3rd. (1999) Modern applications of analytical ultracentrifugation, *Annu. Rev. Biophys. Biomol. Struct. 28*, 75–100.
86. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review., *Protein Sci. 11*, 2067–2079.
87. Rivas, G., Stafford, W., and Minton, A. P. (1999) Characterization of heterologous protein–protein interactions using analytical ultracentrifugation, *Methods 19*, 194–212.
88. Laue, T. M., Anderson, A. L., and Weber, B. J. (1997) Prototype fluorimeter for the XLA/XLI analytical ultracentrifuge, *Proc. SPIE - Inter. Soc. Opt. Eng. 2985*, 196–204.
89. Schuck, P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation, *Biophys. J. 75*, 1503–1512.
90. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling, *Biophys. J. 78*, 1606–1619.

91. Laue, T. (2001) Biophysical studies by ultracentrifugation, *Curr. Opin. Struct. Biol. 11*, 579–583.

92. Ralston, G. (1993) *Introduction to analytical ultracentrifugation*, Beckman Instruments, Inc., Fullerton, CA.

93. McRorie, D. K. and Voelker, P. J. (1993) *Self-associating systems in the analytical ultracentrifuge.*, Beckman Instruments, Inc., Fullerton, CA.

94. Welford, K. (1991) Surface-plasmon polaritons and their uses, *Optical and Quantum Electronics 23*, 1–27.

95. Markey, F. (1999) What is SPR anyway?, *Biacore J. 6*, 14–17.

96. Rich, R. L. and Myszka, D. G. (2000) Advances in surface plasmon resonance biosensor analysis., *Curr. Opin. Biotechnol. 11*, 54–61.

97. Bobst, C. E., Abzalimov, R. R., Houde, D., Kloczewiak, M., Mhatre, R., Berkowitz, S. A., and Kaltashov, I. A. (2008) Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches, *Anal. Chem. 80*, 7473–7481.

98. Philo, J. S. (2009) A critical review of methods for size characterization of non-particulate protein aggregates, *Curr. Pharm. Biotechnol. 10*, 359–372.

99. Hilser, V. J. and Freire, E. (1995) Quantitative-analysis of conformational equilibrium using capillary electrophoresis - applications to protein-folding, *Anal. Biochem. 224*, 465–485.

100. Gavina, J. M. A. and Britz-McKibbin, P. (2007) Protein unfolding and conformational studies by capillary electrophoresis, *Curr. Anal. Chem. 3*, 17–31.

101. Urh, M., Simpson, D., and Zhao, K. (2009) Affinity chromatography: general methods, *Methods Enzymol. 463*, 417–438.

3

OVERVIEW OF BIOLOGICAL MASS SPECTROMETRY

The purpose of this chapter is to provide readers who are less familiar with mass spectrometry with concise background material on modern MS instrumentation and techniques that will be referred to in the later chapters. The idea is to equip the reader with a knowledge base that could later be used to make an informed choice as to the use of a particular instrument or technique for a specific task.

3.1. BASIC PRINCIPLES OF MASS SPECTROMETRY

In a century that has passed since the introduction of the concept of MS by J.J. Thomson (1,2), this technique has become an indispensable analytical tool in many areas of science and technology. This section will briefly review the most important concepts and definitions pertaining to MS. An emphasis will be placed on those features of MS measurements that are particularly relevant for MS analysis of biological macromolecules, as outlined in the following chapters.

The basic principles of mass spectral measurements utilize the ability of electric and magnetic fields to exert influence on charged particles in vacuum:

$$m\ddot{\vec{r}} = q\left(\vec{E} + \left[\dot{\vec{r}} \times \vec{B}\right]\right) \qquad (3\text{-}1\text{-}1)$$

Since the magnitude of each of these interactions is determined by both charge q and mass m of the particle, its trajectory r is universally dependent on the mass-to-charge ratio (m/z). Only charged particles can be manipulated by the electromagnetic fields, hence all neutral molecules must be ionized prior to analysis. Thus, a typical mass spectrometer usually accomplishes two distinct tasks, ionization and mass measurement of the produced ions. According to this view, the mass spectrometer is an analytical device that volatilizes and ionizes molecules, and measures ion abundance as a function of the ionic m/z. In most types of mass spectrometer, such measurements are actually carried out by first separating the ions (either spatially or temporally) according to their m/z, followed by detection of each type of ion (Fig. 3.1). An exception is Fourier transform ion cyclotron resonance (FT ICR), where no physical separation of ions is required prior to ion detection and mass measurement.

Unless the sample is already in the vapor phase, it must be volatilized prior to or concurrently with ionization. The sensitivity of the analysis obviously depends on ionization efficiency, that is, the fraction of the neutral analyte molecules that are converted to ions and transferred to the next segment of the instrument where mass analysis is carried out. Ionization is quite often accompanied by dissociation of the analyte molecules yielding fragment ions, which in many cases can be more abundant than the (quasi) molecular ions. The latter represents the intact analyte molecules M with added charge ($M^{\bullet+}$, MH^+, etc.). Although fragmentation complicates the appearance of the mass spectrum, it often provides useful structural information, a subject discussed in more detail in the following sections of this chapter.

Ions produced in the ionization source are transferred to the mass analyzer, where the m/z ratios are measured. The physical principles of such measurements vary among the many different types of mass analyzers and will be reviewed in Section 3.5. In the remaining part of this section, we will

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

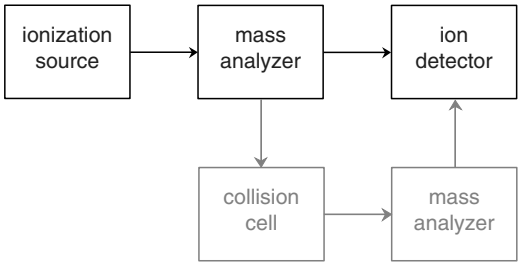


Figure 3.1. A schematic block diagram of a mass spectrometer. Elements shown in gray are used only for tandem MS measurements.

consider some common characteristics and features of mass analysis and their pertinence to the MS of biopolymers.

The total charge q on any ion is a multiple of the elementary (electron or proton) charge e (1.6022×10^{-19} coulomb in SI units):

$$q = \pm ze \tag{3-1-2}$$

where z is always a positive integer. Therefore, the mass-to-charge ratio is always expressed as *m/z*, not *m/q*. The mass of an ion is generally different from that of a neutral molecule. The difference could be quite small (e.g., for ions produced by electron stripping or attachment,* $M^{\cdot+}$ or $M^{\cdot-}$). However, as we will see in the following sections, charge acquisition by biopolymers is almost always accompanied by a substantial gain (or loss) of mass ($[M+nH]^{n+}$, $[M+nH+m\text{Na}]^{(n+m)+}$, $[M-nH+m\text{Na}]^{(n-m)-}$, etc.), which usually cannot be ignored. In this last statement, we make an implicit assumption that a biopolymer under consideration can be assigned a mass, just like any other physical object. As it turns out, the definition of macromolecular mass is not as trivial as it may seem. Since mass measurement is one of the main objectives of MS as an analytical technique, it is necessary to explore the concept of macromolecular mass in greater detail.

3.1.1. Stable Isotopes and Isotopic Distributions

Every molecule has a unique chemical composition; however, there is always some "physical" heterogeneity associated with the presence of stable isotopes (nuclides having the same number of protons, but a different number of neutrons). The existence of stable isotopes was demonstrated experimentally by Francis W. Aston (3), although their existence had been postulated at least a quarter of a century prior to this discovery (for which the Nobel Prize in Chemistry in was awarded in 1922†).

*The electron mass is only 9.1095×10^{-28} g, more than three orders of magnitude below that of a proton (1.6726×10^{-24} g) and neutron (1.6749×10^{-24} g), the building blocks of atomic nuclei.

†Aston's own account of his discovery is available at http://www.nobel.se/chemistry/laureates/1922/aston-lecture.html.

Fractions of "heavy" isotopes of elements that are ubiquitous in biological molecules (C, N, H, O, P, etc.) do not usually exceed 1%‡ (see Table 3.1). Nevertheless, if a molecule contains a large number of atoms of a certain element, contributions of heavy isotopes can become very significant. We will illustrate this using carbon, an element that has two stable, naturally occurring isotopes, ^{12}C (98.9%) and ^{13}C (1.1%), and buckminsterfullerene, a C_{60} molecule. The probability that every carbon atom in C_{60} is represented by the "light" isotope can be calculated as $P_o = (0.989)^{60} = 0.515$, meaning that almost one-half of all buckminsterfullerene molecules contain at least one ^{13}C atom. The probability that k and only k out of 60 carbon atoms are ^{13}C can be easily calculated as:

$$P_k = C_k^{60} \cdot (0.011)^k \cdot (0.989)^{60-k} \tag{3-1-3}$$

where C_k^{60} is a binomial coefficient (number of possible combinations of k elements out of 60). The graph of P_k as a function of mass or, more precisely, number of ^{13}C atoms (Fig. 3.2) represents an *isotopic distribution* of buckminsterfullerene. A trivial expansion of eq. 3-1-3 gives an isotopic distribution for a hypothetical molecule built of n atoms A with two stable isotopes, ^{X}A (relative abundance p_0) and ^{X+1}A (relative abundance $p_1 = 1 - p_0$):

$$P_k = C_k^n \cdot (p_1)^k \cdot (p_0)^{n-k} \tag{3-1-4}$$

It is easy to show that eq. 3-1-4 can be expanded to include more than two elements, for example, an isotopic distribution for a molecule A_nB_m can be expressed as a convolution of two binomial distributions:

$$P_k = \sum_{i,j} C_i^n \cdot C_j^m \cdot \left(p_1^{A}\right)^i \cdot \left(p_0^{A}\right)^{n-i} \cdot \left(p_1^{B}\right)^j \cdot \left(p_0^{B}\right)^{m-j} \delta_{k,i+j} \tag{3-1-5}$$

where $p_0{}^{A}$, $p_1{}^{A}$, $p_0{}^{B}$, and $p_1{}^{B}$ represent the relative abundance of ^{X}A, ^{X+1}A, ^{Y}B, and ^{Y+1}B, respectively, and $\delta_{k,i+j}$ is a Kronecker symbol ($\delta_{p,q} = 1$ if $p = q$, and 0 if $p \neq q$), or simply

$$P_k = \sum_{i \le k} C_i^n \cdot C_{k-i}^m \cdot \left(p_1^{A}\right)^i \cdot \left(p_0^{A}\right)^{n-i} \cdot \left(p_1^{B}\right)^{k-i} \cdot \left(p_0^{B}\right)^{m-k+i} \tag{3-1-6}$$

Another, more intuitive presentation of eq. 3-1-5 uses a product of the two binomials, $(p_0{}^{A} + p_1{}^{A})^n (p_0{}^{B} + p_1{}^{B})^m$, whose expansion gives the relative abundance P_k of the isotopic species whose mass exceeds that of $^{X}A^{Y}B$ by k

‡Sulfur is a notable exception, with ^{33}S and ^{34}S together accounting for almost 5% of stable sulfur isotopes.

TABLE 3.1. Isotope Abundance and Accurate Masses for Selected Elements Based on Compilations[a–d]

Element	Isotope	Accurate Mass, *u* (STD, μ*u*)	Natural Abundance
Hydrogen	^{1}H	1.007 825 031 90 (0.000 57)[a]	0.999 844 26(5)[b]
	^{2}H (D)	2.014 101 777 95 (0.000 62)[a]	0.000 155 74(5)[b]
Carbon[b]	^{12}C	12.000 000 000 000	0.988 944(28)
	^{13}C	13.003 354 838 3 (0.0049)[a]	0.011 056(28)[b]
Nitrogen	^{14}N	14.003 074 007 4 (0.0018)[a]	0.996 337(4)[b]
	^{15}N	15.000 108 973 (0.012)[a]	0.003 663(4)[a]
Oxygen	^{16}O	15.994 914 622 3 (0.0025)[a]	0.997 6206(5)[a]
	^{17}O	16.999 131 50 (0.22)[b]	0.000 3790(9)[b]
	^{18}O	17.999 160 4 (0.9)[b]	0.002 0004(5)[b]
Fluorine	^{19}F	18.998 403 2 (0.5)[c]	1[e]
Sodium	^{23}Na	22.989 770 (2)[c]	1[e]
Magnesium[b]	^{24}Mg	23.985 041 87 (0.26)	0.789 92(25)
	^{25}Mg	24.985 837 00 (0.26)	0.100 03(9)
	^{26}Mg	25.982 593 00 (0.26)	0.110 05(19)
Aluminum	^{27}Al	26.981 538 (2)[c]	1[e]
Silicon[b]	^{28}Si	27.976 926 49 (0.22)	0.922 223(9)
	^{29}Si	28.976 494 68 (0.22)	0.046 853(6)
	^{30}Si	29.973 770 18 (0.22)	0.030 924(7)
Phosphorus	^{31}P	30.973 761 (2)[c]	1[e]
Sulfur[b]	^{32}S	31.972 070 73 (0.15)	0.950 3957(90)
	^{33}S	32.971 458 54 (0.15)	0.007 4865(12)
	^{34}S	33.967 866 87 (0.14)	0.041 9719(87)
	^{36}S$^{(\beta-)}$	35.967 080 88 (0.25)	0.000 1459(21)
Chlorine[b]	^{35}Cl	34.968 852 71 (0.04)	0.757 79(46)
	^{37}Cl	36.965 902 60 (0.05)	0.242 21(46)
Potassium	^{39}K	38.963 706 9 (0.3)[a]	0.932 581 (44)[d]
	^{40}K$^{(\beta-)}$	39.963 998 67 (0.29)[a]	0.000 117 (1)[d]
	^{41}K	40.961 825 97 (0.28)[a]	0.067 302 (44)[d]
Calcium[b]	^{40}Ca	39.962 591 2 (0.3)	0.969 41(6)
	^{42}Ca	41.958 618 3 (0.4)	0.006 47(3)
	^{43}Ca	42.958 766 8 (0.5)	0.001 35(2)
	^{44}Ca	43.955 481 1 (0.9)	0.020 86(4)
	^{46}Ca$^{(\beta-)}$	45.953 692 7 (2.5)	0.000 04(1)
	^{48}Ca$^{(\beta-)}$	47.952 533 (4)	0.001 87(1)
Manganese	^{55}Mn	54.938 049 (9)[c]	1[e]
Iron[b]	^{54}Fe	53.939 614 7 (1.4)	0.058 45(23)
	^{56}Fe	55.934 941 8 (1.5)	0.917 54(24)
	^{57}Fe	56.935 398 3 (1.5)	0.021 191(65)
	^{58}Fe$^{(\beta-)}$	57.933 280 1 (1.5)	0.002 819(27)
Nickel	^{58}Ni	57.935 347 7 (1.6)[a]	0.680769 (89)[d]
	^{60}Ni	59.930 790 3 (1.5)[a]	0.262231 (77)[d]
	^{61}Ni	60.931 060 1 (1.5)[a]	0.011399 (6)[d]
	^{62}Ni	61.928 348 4 (1.5)[a]	0.036345 (17)[d]
	^{64}Ni$^{(\beta-)}$	63.927 969 2 (1.6)[a]	0.009256 (9)[d]
Copper[b]	^{63}Cu	62.929 600 7 (1.5)	0.691 74(20)
	^{65}Cu	64.927 793 8 (1.9)	0.308 26(20)
Zinc[b]	^{64}Zn	63.929 146 1 (1.8)	0.4863(20)
	^{66}Zn	65.926 036 4 (1.7)	0.2790(9)
	^{67}Zn	66.927 130 5 (1.7)	0.0410(4)
	^{68}Zn	67.924 847 3 (1.7)	0.1875(17)
	^{70}Zn$^{(\beta-)}$	69.925 325 (4)	0.0062(1)
Gallium	^{69}Ga	68.925 581 (3)[a]	0.60108 (9)[d]
	^{71}Ga	70.924 707 3 (2.0)[a]	0.39892 (9)[d]
Arsenic	^{75}As	74.921 60 (20)[c]	1[e]
Selenium[b]	^{74}Se	73.922 476 7 (1.6)	0.008 89(3)

TABLE 3.1. (*Continued*)

Element	Isotope	Accurate Mass, u (STD, μu)	Natural Abundance
	^{76}Se	75.919 214 3 (1.6)	0.093 66(18)
	^{77}Se	76.919 914 8 (1.6)	0.076 35(10)
	^{78}Se$^{(\beta-)}$	77.917 309 7 (1.6)	0.237 72(20)
	^{80}Se$^{(\beta-)}$	79.916 522 1 (2.0)	0.496 07(17)
	^{82}Se$^{(\beta-)}$	81.916 700 3 (2.2)	0.087 31(10)
Bromine	^{79}Br	78.918 337 9 (2.0)[a]	0.5069 (7)[d]
	^{81}Br	80.916 291 (3)[a]	0.4931 (7)[d]
Ruthenium	^{96}Ru	95.907 604 (9)[a]	0.0554 (14)[d]
	^{98}Ru	97.905 287 (7)[a]	0.0187 (3)[d]
	^{99}Ru	98.905 938 5 (2.2)[a]	0.1276 (14)[d]
	^{100}Ru	99.904 218 9 (2.2)[a]	0.1260 (7)[d]
	^{101}Ru	100.905 581 5 (2.2)[a]	0.1706 (2)[d]
	^{102}Ru$^{(\beta-)}$	101.904 348 8 (2.2)[a]	0.3155 (14)[d]
	^{104}Ru$^{(\beta-)}$	103.905 430 (4)[a]	0.1862 (27)[d]
Cadmium	^{106}Cd	105.906 458 (6)[a]	0.0125 (6)[d]
	^{108}Cd	107.904 183 (6)[a]	0.0089 (3)[d]
	^{110}Cd	109.903 006 (3)[a]	0.1249 (18)[d]
	^{111}Cd	110.904 182 (3)[a]	0.1280 (12)[d]
	^{112}Cd$^{(\beta-)}$	111.902 757 7 (3.0)[a]	0.2413 (21)[d]
	^{113}Cd$^{(\beta-)}$	112.904 401 4 (3.0)[a]	0.1222 (12)[d]
	^{114}Cd$^{(\beta-)}$	113.903 358 6 (3.0)[a]	0.2873 (42)[d]
	^{116}Cd$^{(\beta-)}$	115.904 756 (3)[a]	0.0749 (18)[d]
Indium	^{113}In	112.904 062 (4)[a]	0.0429 (5)[d]
	^{115}In$^{(\beta-)}$	114.903 879 (4)[a]	0.9571 (5)[d]
Iodine	^{127}I	126.904 468 (4)[a]	1[e]
Gadolinium	^{152}Gd	151.919 789 (3)[a]	0.0020 (1)[d]
	^{154}Gd$^{(\beta-)}$	153.920 862 (3)[a]	0.0218 (3)[d]
	^{155}Gd	154.922 619 (3)[a]	0.1480 (12)[d]
	^{156}Gd$^{(\beta-)}$	155.922 120 (3)[a]	0.2047 (9)[d]
	^{157}Gd	156.923 957 (3)[a]	0.1565 (2)[d]
	^{158}Gd$^{(\beta-)}$	157.924 101 (3)[a]	0.2484 (7)[d]
	^{160}Gd$^{(\beta-)}$	159.927 051 (3)[a]	0.2186 (19)[d]
Mercury	^{196}Hg	195.965 814 (4)[a]	0.0015 (1)[d]
	^{198}Hg	197.966 752 (3)[a]	0.0997 (20)[d]
	^{199}Hg	198.968 262 (3)[a]	0.1687 (22)[d]
	^{200}Hg	199.968 309 (3)[a]	0.2310 (19)[d]
	^{201}Hg	200.970 285 (3)[a]	0.1318 (9)[d]
	^{202}Hg	201.970 625 (3)[a]	0.2986 (26)[d]
	^{204}Hg$^{(\beta-)}$	203.973 475 (3)[a]	0.0687 (15)[d]
Bismuth	^{209}Bi	208.980 38 (20)[a]	1[e]

[a] Reference (4).
[b] Reference (5).
[c] Reference (6).
[d] Reference (7).
[e] Single stable isotope for this element.

mass units. Further expansion to a general case of a polyatomic molecule $A_nB_m\ldots Z_q$ is relatively straightforward:

$$(p_0^A+p_1^A+p_2^A+\cdots)^n\cdot(p_0^B+p_1^B+p_2^B+\cdots)^m\cdots (p_0^Z+p_1^Z+p_2^Z+\cdots)^q \quad (3\text{-}1\text{-}7)$$

Despite its analytical elegance, Eq. 3-1-7 is computationally demanding and is rarely used for calculating the isotopic distributions in a straightforward manner. Multiple algorithms have been developed over the last couple of decades that use various robust schemes to calculate isotopic distributions of large polyatomic molecules with high accuracy (8–13). Isotopic distributions of many elements (metals in particular) display some rather characteristic patterns, which sometimes provide unique "isotopic signatures" within small molecules (Fig. 3.3). Isotopic distributions of most

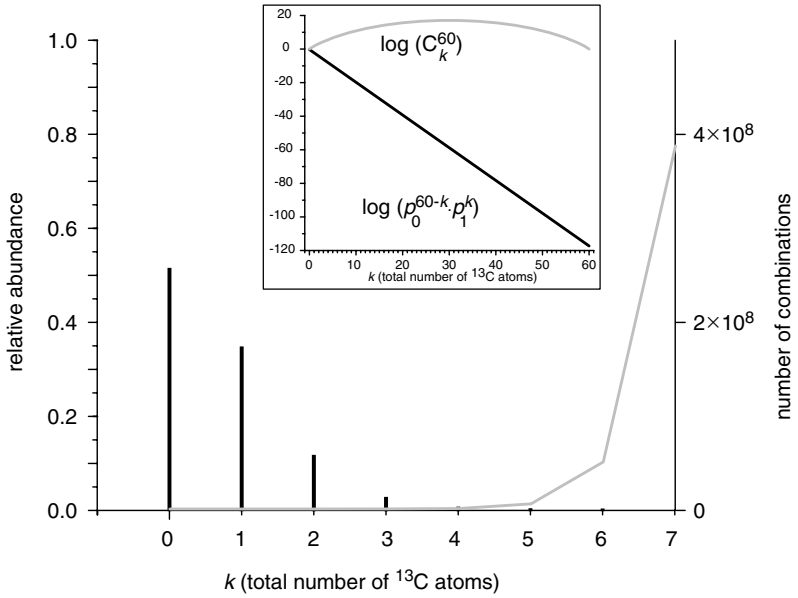


Figure 3.2. Isotopic distribution of a buckminsterfullerene (C_{60}) molecule (shown as a bar diagram). The inset shows logarithms of a number of combinations of k atoms out of 60 (gray trace) and a probability that in each such combination k atoms are ^{13}C isotopes (black trace).

biopolymers are not as telling (as far as the elemental make-up is concerned). Instead, the large number of atoms present in a "typical" biomolecule gives rise to a convoluted isotopic distribution whose width increases as the size of the macromolecule increases (Fig. 3.4).

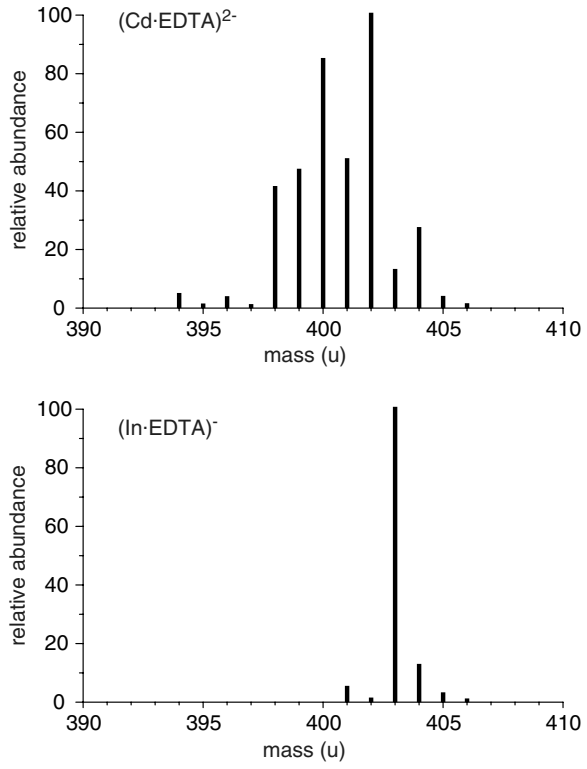


Figure 3.3. Isotopic signatures of metal ion–organic acid complexes: isotopic distributions of $(Cd \cdot EDTA)^{2-}$ and $(In \cdot EDTA)^{-}$, where EDTA=ethylenediaminetetraacetic acid.

Most commercial mass spectrometers have built-in software that calculates isotopic distributions.* The data input is often available in two formats: general (using a molecular formula $A_nB_m \ldots Z_q$) or peptide - and protein-oriented (using an amino acid composition $Aaa_nBbb_m \ldots Zzz_q$ or simply an amino acid sequence). Most of the web-based programs currently calculate isotopic distributions assuming the natural abundance of stable isotopes for each element. Calculations of isotopic distributions for proteins expressed in a medium that is enriched with (or depleted of) a certain isotope(s) would require a more flexible algorithm.

Strictly speaking, in each of the calculated isotopic distributions shown in Figure 3.4, only the monoisotopic peak is truly a single peak, with the rest of the peaks in the cluster being composite peaks representing two or more isobaric species. For example, the monoisotopic peak of ubiquitin shown in Figure 3.4 corresponds to an ionic species $(^{12}C_{378}{}^{1}H_{631}{}^{14}N_{105}{}^{16}O_{118}{}^{32}S)^{+}$, while the next peak represents five different species, which are usually referred to as *isobaric species* (or *isobars)*. These five isobars $(^{12}C_{377}{}^{13}C^{1}H_{631}{}^{14}N_{105}{}^{16}O_{118}{}^{32}S)^{+}$, $(^{12}C_{378}{}^{1}H_{630}{}^{2}H^{14}N_{105}{}^{16}O_{118}{}^{32}S)^{+}$, $(^{12}C_{378}{}^{1}H_{631}{}^{14}N_{104}{}^{15}N^{16}O_{118}{}^{32}S)^{+}$, $(^{12}C_{378}{}^{1}H_{631}{}^{14}N_{105}{}^{16}O_{117}{}^{17}O^{32}S)^{+}$, and $(^{12}C_{378}{}^{1}H_{631}{}^{14}N_{105}{}^{16}O_{118}{}^{33}S)^{+}$, have slightly different masses due to unequal divergence of ^{13}C, ^{2}H, ^{15}N, ^{17}O, and ^{33}S from the nearest

*A large array of such programs is also available on the Internet, such as the "MS Isotope" routine available at the popular Protein Prospector website at http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msisotope).

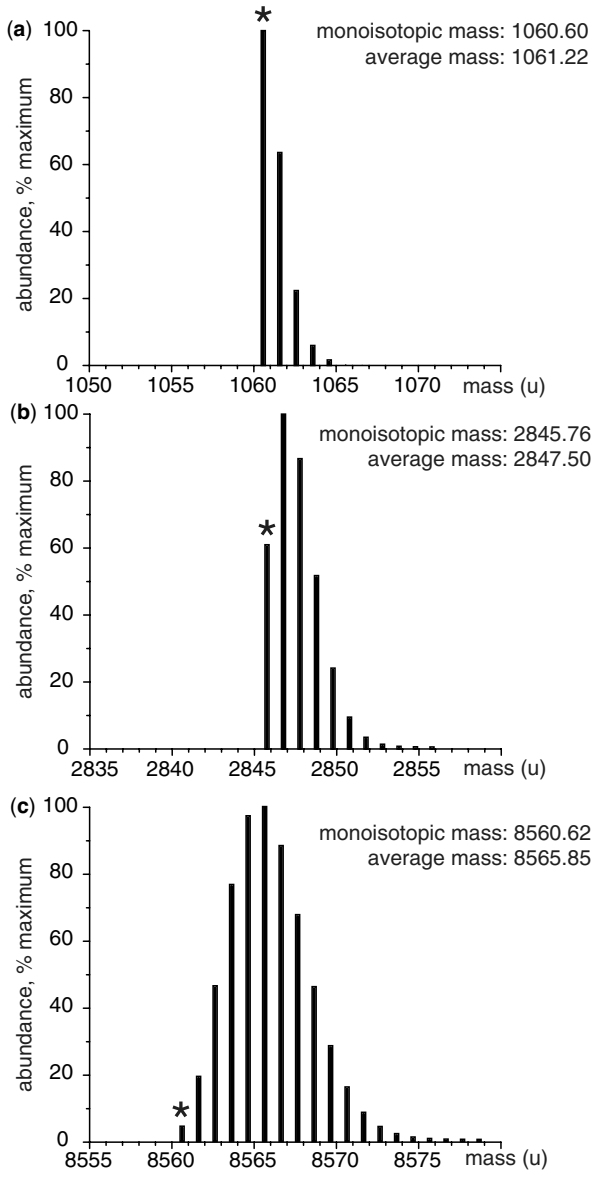


Figure 3.4. Calculated isotopic distributions for polypeptides bradykinin (*a*), melittin (*b*) and ubiquitin (*c*). Monoisotopic peaks are marked with asterisks.

integer number (see Table 3.1). Although the isobars of smaller peptides can be resolved in some cases (14), the present level of mass measurement technology does not generally allow such distinction to be made for larger (>1 kDa) peptides and proteins. Therefore, hereafter we make no distinction between the isobars, and the "composite" nature of "isotopic peaks" will generally be ignored.

3.1.2. Macromolecular Mass: Terms and Definitions

The notion of *molecular mass*, as it is used in MS, is closely related to, but not necessarily the same as, the familiar concept of *molecular weight,* a sum of the atomic weights of all atoms in the molecule. Molecular mass is measured in unified atomic mass units (u), defined by International Union of Pure and Applied Chemistry (IUPAC) as one-twelfth of the mass of a carbon-12 (^{12}C) atom in its ground state, u $\approx 1.6605402(10) \times 10^{-27}$ kg. The *atomic weight* of an element is a weighted average of the atomic masses of the different isotopes, therefore the isotopic make-up is implicitly included in the definition. Since the ionic mass measured by MS is not necessarily averaged across the entire isotopic content, several definitions of molecular mass are currently in circulation. The definitions of the molecular mass vary based on how they account for contributions from different isotopes. The *nominal mass* is calculated using masses of lightest isotopes for each element rounded to the nearest integer, although some texts use the most abundant (instead of the lightest) isotope to define the nominal mass.

The *monoisotopic mass* is calculated in a similar fashion, but the isotopic masses are no longer rounded, that is, the nuclear mass defect is accounted for. For peptides whose molecular weight exceeds $\sim$ 2 kDa, the *most abundant mass* (i.e., mass corresponding to the ionic peak of highest intensity in the isotopic cluster) no longer coincides with the monoisotopic mass (Fig. 3.4). Finally, the *average mass* is calculated based on the entire isotopic distribution and is closely related to the *molecular weight* as used elsewhere in chemistry and related disciplines. The average mass is usually very close to the most abundant mass (typically within 1 u). As the number of atoms comprising the molecule increases, so does the difference between the *monoisotopic* and *average* masses. At the same time, the relative abundance of the monoisotopic peak continues to decrease (Fig. 3.4), and it becomes practically undetectable even for biopolymers of a modest (> 10 kDa) size.

3.2. METHODS OF PRODUCING BIOMOLECULAR IONS

3.2.1. Macromolecular Ion Desorption Techniques: General Considerations

Despite its early success as an analytical technique, MS had made almost no incursions into the field of intact biopolymer analysis prior to the 1970s. The "classical" ionization techniques (e.g., electron impact and chemical ionization) were not suited to handle biological macromolecules. The vapor pressure of biopolymers is negligible, and their delicate nature prevented using high temperatures as a means to enhance evaporation. Although this difficulty was circumvented in several cases by the chemical derivatization of polar biomolecules (aiming at increasing their vapor pressure), examples of using "classical" MS to analyze even the simplest biomolecules remained very few.

Mass spectral analysis of biomolecules only became feasible with the advent of ion desorption techniques, which initially included field desorption [reviewed in (15)] and plasma desorption (16). The introduction of fast atom bombardment (FAB) in the early 1980s (17) had perhaps even greater impact on the development of biomolecular MS. Unlike plasma desorption, FAB sources did not require special types of mass analyzers and were much easier to handle. While plasma desorption and field desorption–ionization sources are no longer used widely, FAB remains in use (18). However, its application in biophysical experiments is usually limited, and we will not consider it here in much detail. An interested reader is referred to several excellent reviews on the subject (19,20), as well as comprehensive MS texts (21,22). The two ionization techniques that are most relevant to our discussion are electrospray ionization mass spectrometry (ESI MS) and matrix-assisted laser desorption–ionization (MALDI).

3.2.2. Electrospray Ionization

Brief Historical Remarks. The advent of ESI MS in the mid-1980s (23) provided a means to observe spectra of intact proteins with no apparent mass limitation, a discovery honored with a Nobel Prize in Chemistry to John Fenn in 2002 (24). The history of developing this technique is quite fascinating. The ESI process had been in use in various fields of science and technology (including MS) for several decades prior to its direct application to biopolymer analysis. The "electrification of liquid droplets produced by spraying, bubbling, and similar methods" (25) had captured the attention of a number of researchers as early as the end of the nineteenth and beginning of the twentieth century, with the initial theoretical exploration of the ESI process carried out by Lord Rayleigh in 1882 (26). Indeed, studies of some of the phenomena related to ESI can be traced several hundred years further back into the past (27).

The motivation for earlier studies of ion production and fate is mostly due to the importance of such processes in atmospheric science. The realization that ESI held great promise and potential to become a means of producing macromolecular ions started fueling the interest of the MS community in the late 1960s (28). Malcolm Dole, a pioneer of ESI MS, wrote in 1968: In considering how it might be possible to obtain gas phase intact ionized macromolecules for mass analysis in a mass spectrometer, the idea occurred to one of us to electrospray dilute solutions of macromolecules into air or other suitable gas at atmospheric pressure and to sample the air for macroions by means of a supersonic probe. By allowing a dilute polymer solution containing a volatile solvent to flow out of a tip of a hypodermic needle electrostatically charged. . ., a spray of finely divided and electrically charged droplets is produced. On evaporation of the solvent the charged droplets should become electrically unstable and break down into smaller droplets . . . [eventually resulting in the formation of] drops containing only one macromolecule per drop. Finally, on complete evaporation of solvent from these drops, . . . the gas phase macroions would result (29) (see Fig. A1). In fact, the first application of ESI to observe macromolecules and measure their molecular weight was reported in 1964 (30). Detection, though, was done using optical (not mass spectrometric) means. The subsequent efforts of Dole and co-workers (31,32) focused on producing polymer ions in the gas phase. Similar ideas were later used by Iribarne, Thomson and co-workers (33,34) to develop an ionization method capable of dealing with polar and labile analytes, which they termed atmospheric pressure ion evaporation. The technique was shown to be quite capable of producing molecular ions for a range of small polar organic molecules, including amino acids (33,35,36) and adenosine triphosphate (ATP) (33). Concurrently, Aleksandrov et al. used ESI MS for peptide analysis (37,38) and reported controlled peptide ion dissociation in the interface region yielding structurally diagnostic fragment ions, a phenomenon that later would be referred to as *cone fragmentation* or *nozzle-skimmer fragmentation* (35). Finally, an ESI source was interfaced with high-performance liquid chromatography (HPLC) (39,40). However, as Vestal (41) recalls, most of this work was largely ignored by the majority of practitioners of biological MS.

A breakthrough occurred in 1988 when Fenn's group demonstrated ESI mass spectra of intact proteins (42) and synthetic polymers (43) and presented their findings at the 36th Annual conference of the American Society for Mass Spectrometry. Finally, Dole's bold suggestion that it would be possible to "*obtain gas phase intact ionized macromolecules for mass analysis in a mass spectrometer"* was shown to be true. An interested reader may find more information on this subject by listening to Fenn's lecture "Electrospray wings for molecular elephants", presented in December 2002 upon his acceptance of the Nobel Prize in Chemistry (video stream at http://www.nobel.se/chemistry/laureates/2002/fenn-lecture.html).

Macro-Ions in ESI: Multiple Charging. Electrospray ionization is a convoluted process that involves several steps, each having a profound effect on the outcome of the measurements (Fig. 3.5). Many of these effects are particularly important in biophysical measurements and must be taken into account in order to avoid erroneous interpretation of the experimental data. A detailed discussion of various physical processes involved in ESI is presented in the Appendix.

A very distinct feature of ESI is the accumulation of multiple charges on a single analyte molecule during ionization; an important implication of this is the appearance of multiple peaks in the mass spectrum corresponding to a

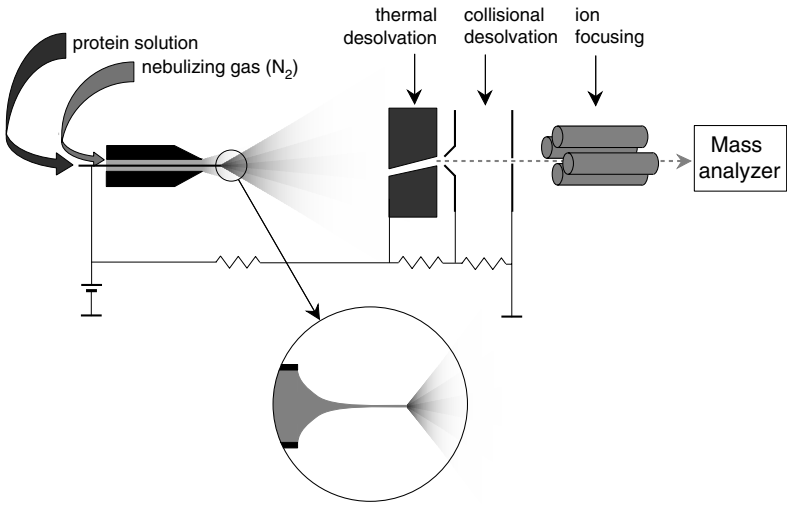


Figure 3.5. A schematic representation of an ESI MS interface.

single analyte (Fig. 3.6*a*). The position of each peak (*m*/*z* value) would be determined by the mass of a macromolecule, *M*, and the number of accommodated charges. There are several different charge carriers that are commonly associated with an ESI process, the most ubiquitous being a proton and alkali metal cations (particularly Na^+ and K^+). Given such heterogeneity in the charges accommodated by a macromolecular ion, its *m*/*z* value can be generally calculated as:

$$\frac{m}{z} = \frac{M + \sum_i n_i m_i}{\sum_i n_i} \qquad (3\text{-}2\text{-}1)$$

where m_i is a mass associated with a specific carrier, and n_i is a number of charges of this type associated with the macromolecule. If the charging is mostly due to protonation, then Eq. 3-2-1 is simplified to

$$\left(\frac{m}{z}\right)_{n+} = \frac{M+n}{n} \qquad (3\text{-}2\text{-}2)$$

Mass spectra of even relatively small protein ions typically contain several peaks corresponding to different charge states. Therefore, calculating the mass *M* based on a series of *m*/*z* values of ion peaks corresponding to incrementally increasing (or decreasing) charge states is an overdetermined problem. Such calculations are easy and very straightforward (inset in Fig. 3.6). Spectral interpretation becomes a bit more difficult if an ESI mass spectrum contains contributions from several different analytes. Multiple deconvolution procedures have been developed to address this problem (44–46), and most modern ESI MS data systems contain built-in deconvolution routines. High-resolution MS provides an alternative to deconvolution, as the charge state of each macromolecular ion can be deduced directly from the mass spectrum based on the distance between the adjacent isotopic peaks (Fig. 3.6*b*). The most sophisticated (and most accurate) methods of mass analysis of complex mixtures make use of

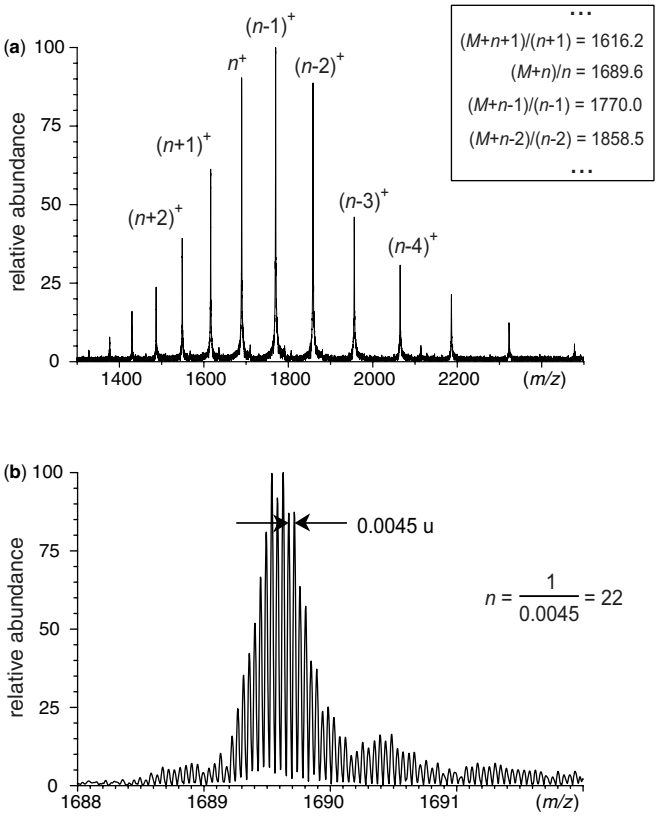


Figure 3.6. The ESI mass spectrum of a 37 kDa protein human serum transferrin N-lobe (*h*Tf/2N) in $H_2O/CH_3OH/CH_3COOH$ (49:49:2, *v*:*v*:*v*) showing a series of multiply charged protein ion peaks (*a*) and an expanded view of the isotopic cluster of a 22+ ion of *h*Tf/2N (*b*). [Courtesy of Mingxuan Zhang (presently at Biogen IDEC, Cambridge, MA).]

both deconvolution procedures and high-resolution measurements (47–49).

The presence of other modes of charging a macromolecule (besides protonation) usually degrades the quality of ESI MS data, as the deconvoluted spectra in this case contain artifact peaks (e.g., corresponding to Na^+ and K^+ adducts). Although not a significant annoyance by itself, it usually leads to a diminished S/N ratio due to a distribution of the total ionic signal among multiple charge states and adduct species, and in some extreme cases to an inferior accuracy of mass measurements. Other types of adducts that are encountered in ESI MS are ammonium cations and their derivatives. Adducts formed by association of ubiquitous anions (acetate, formate, etc.) with positively charged macromolecular species are also common. The extent of adduct ion formation can often be limited by using heat-induced desolvation (via elevating the temperature in the ESI interface) and/or employing carefully controlled collision-induced dissociation of the adduct ions. Each of these processes can greatly reduce the extent of biopolymer ion complexation with anions, while Na^+ and K^+ adducts remain largely unaffected. Excessive collisional activation brings about apparent reduction of the macromolecular ion charge state (the "charge stripping phenomenon") and may lead to the dissociation of covalent bonds, a process that will be considered in detail in Section 3.4.

The formation of macromolecular ionic species in the negative ion mode usually proceeds via deprotonation of acidic groups. Deconvolution of negative ion ESI mass spectra is relatively straightforward, although one has to remember that acquisition of the negative charge by biopolymers most often occurs through deprotonation (mass loss), so the *m/z* values of protein ions peaks are

$$\left(\frac{m}{z}\right)_{n-} = \frac{M-n}{n} \qquad (3\text{-}2\text{-}3)$$

Most classes of biological polymers are amenable to ESI MS analysis, including proteins, oligonucleotides, and carbohydrates. Proteins are usually analyzed in the positive ion mode, while oligonucleotides tend to produce a better signal in the negative ion mode. Alkali metal adduct ion formation often becomes a significant problem for larger oligonucleotides, while Na^+ (or K^+) complexation is often used intentionally as a means to ionize neutral carbohydrates lacking basic and acidic groups.

Applications of ESI MS for biopolymer analysis constitute a vast and still rapidly expanding field. Unfortunately, obvious space limitations do not allow us to explore this field beyond the subjects relevant to biophysics. An interested reader is referred to a couple of excellent books on the subject that contain diverse examples of ESI MS applications to various problems in the life sciences in general (50,51). We will, however, briefly consider one very popular modification of ESI MS that is particularly important for some of the biophysical problems to be discussed in the following chapters of the book.

Nanospray Ionization. In traditional ESI MS, the analyte solution is continuously supplied to the ion source at a constant flow rate. Although the flow rate can vary greatly, it rarely goes below the microliter per minute (μL/min) level, as it causes various spray instabilities when using conventional ESI sources. Wilm and Mann (52) pointed out that achieving very low flow rates could be beneficial in many ways. Since the diameter of the emitted droplets is determined by the liquid flow rate through the capillary, only very small droplets are formed when the flow rate is dramatically reduced. Large surface-to-volume ratio results in facile droplet evaporation and significant improvement of the efficiency of ion formation. Of course, another benefit is a dramatic extension of the analysis time, given the sample volume remains constant. Realization of these ideas led to development of the nano-electrospray ionization (nano-ESI) (53). The very low flow rate is realized by loading a small volume (typically 1 μL) of sample solution into a narrow bore (orifice 1–2 μm) metal-coated capillary. Liquid flow is actually induced by applying high voltage to the capillary tip, (i.e., the solution is drawn from the capillary electrodynamically without the use of a conventional syringe pump). The resulting flow rates are typically on the order of 20–40 nL/min and are quite stable (the small orifice prevents the formation of multiple Taylor cones at the tip of the capillary). In addition to increased sensitivity as compared to conventional ESI, nano-ESI has another important advantage. Several studies have indicated that it has much higher salt tolerance, at least an order of magnitude exceeding that of conventional ESI (54,55). This finding is explained in terms of the lower size and higher charge density of droplets emitted in nano-ESI, which result in early fission events without extensive solvent evaporation, which would otherwise lead to a significant increase in salt ion concentration prior to fission.

3.2.3. Matrix Assisted Laser Desorption Ionization (MALDI)

Electrospray ionization is an example of a class of ionization methods that produce macromolecular ions directly from the solution bulk (other "bulk ionization" methods have met with less success when applied to biopolymers). Desorption from solid surfaces provides another opportunity to transfer analyte molecules to the gas phase. Perhaps the simplest method, heating to vaporize the sample, has become the *de facto* standard for analysis of small organic molecules, where ionization is achieved by passing the gaseous analyte through an electron beam. The harsh

conditions of heating and electron impact, however, generally lead to extensive fragmentation, and this method is only practical for low molecular weight (<800 Da) and relatively nonpolar analytes. A reasonable alternative to the indiscriminate heating of the sample is the utilization of high-energy projectiles that serve as "external agents" providing very localized heating to the sample. Such rapid and local heating would result in a near-adiabatic expansion of microscopically small portions of the sample, leading to their expulsion from the surface to the gas phase.

This principle is utilized in FAB ionization, which is mentioned in Section 3.2.1, where fast atoms or ions are utilized as high-energy projectiles. Rapid and highly localized heating of the sample can also be achieved by using photons instead of heavy particles. Utilization of laser beams allows light energy of high intensity to be focused on a small area, facilitating desorption and ionization of species from either a solid or liquid sample. The rate of energy transfer, dependent on the laser fluence, determines whether vaporization is favored over decomposition of the analyte. If sample heating upon irradiation is sufficiently rapid, then it becomes possible, in many cases, to desorb intact species before decomposition can occur. This technique, known as laser desorption (or laser ablation) MS (56) generally required postionization of the analyte molecules in the gas phase. A major limitation of laser desorption is the inadequate efficiency of energy transfer to the irradiated sample, without which a facile desorption of intact macromolecules cannot be achieved. This difficulty was solved by the utilization of chromophores as matrices that facilitate the energy transfer and effectively shield the analyte molecules from radiation damage (Fig. 3.7). This technique, presently known as MALDI was developed simultaneously by Tanaka et al. (57) and Karas and Hillenkamp (58), an invention for which the Nobel Prize in Chemistry was awarded in 2002 (59). Most current MALDI schemes use aryl-based acids (e.g., nicotinic acid) as UV-absorbing matrices, an approach pioneered by Karas and Hillenkamp (58). Many organic compounds with conjugated double bonds absorb ultraviolet (UV) light in the 250–370-nm range (Fig. 3.8), hence the popularity of relatively inexpensive nitrogen lasers in MALDI MS with emitting wavelength of 337 nm (another popular choice is the Nd:YAG laser emitting at 355 nm). The approach pioneered by Tanaka (utilization of small light-absorbing particles to assist ion desorption and ionization) is also beginning to enjoy a renaissance with the emergence and rapid proliferation of nanoparticles in MS analyses (60).

Soon after the introduction of UV–MALDI MS, several groups started experimenting with infrared (IR) lasers as a means of macromolecule desorption from the solid surface (61–63). A particularly intriguing aspect of IR–MALDI is the possibility of utilizing frozen water as a matrix (water has a strong absorption band near 3 μm due to the O—H stretching mode), although a range of other matrices can also be utilized, including glycerol and urea. Using water and glycerol as MALDI matrices may open up several interesting opportunities. For example, it may allow biological macromolecules to be kept in their native environments prior to mass analysis (whereas most UV matrices denature proteins), or interface MALDI MS more readily with HPLC (64).

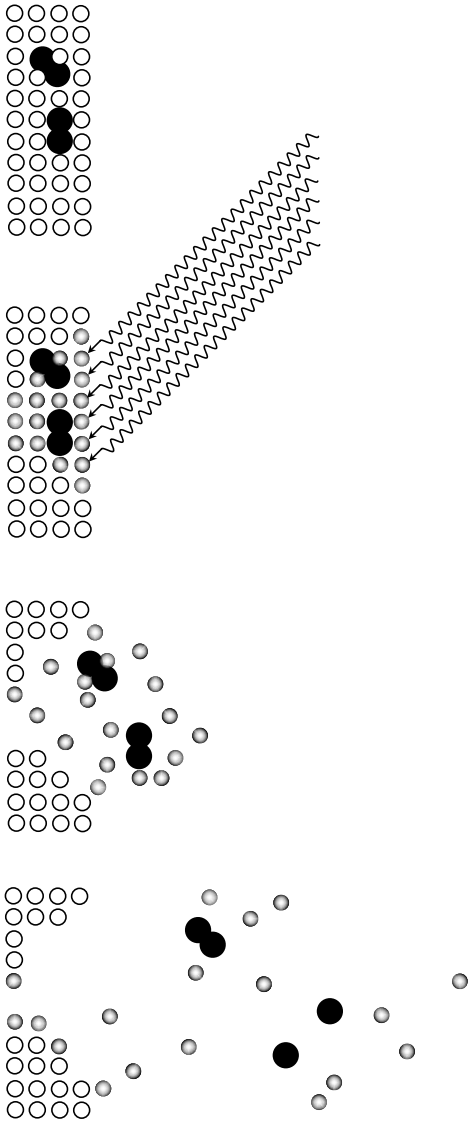


Figure 3.7. A schematic representation of the MALDI process. White circles represent matrix molecules packed in a crystal; embedded analyte molecules are shown in black. Gray circles represent photoexcited matrix molecules.

Macromolecular ions produced by MALDI can also carry multiple charges; however, the extent of protonation is significantly below that achieved with ESI (Fig. 3.9). This requires the utilization of mass analyzers with an extended *m/z* range [e.g., time-of-flight (TOF)], although analyzers

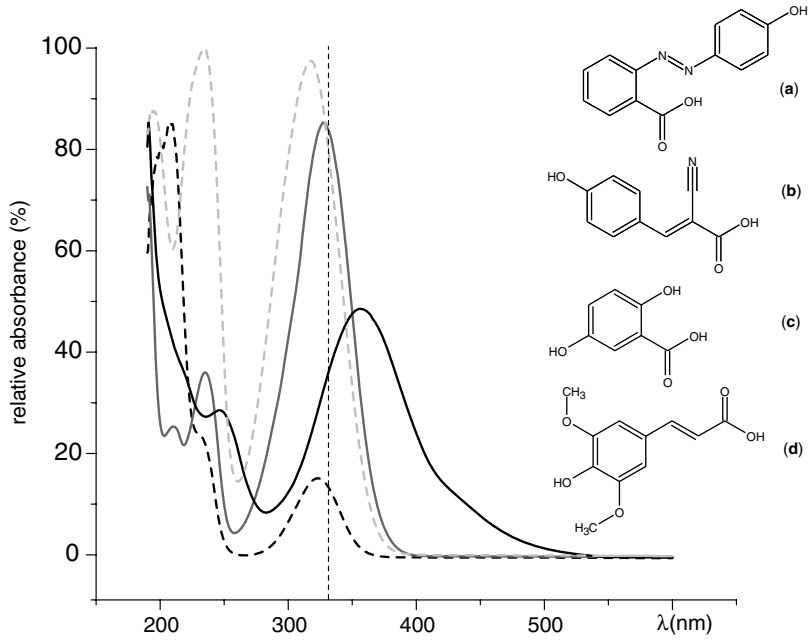


Figure 3.8. The UV–Vis absorption spectra and chemical structures of popular MALDI matrices: (*a*, solid black trace) 2-(4-hydroxyphenylazo)–benzoic acid (HABA), (*b*, solid gray trace) α-cyano-4-hydroxy cinnamic acid (αCHCA), (*c*, dashed black trace) 2,5-dihydroxybenzoic acid (DHBA), and (*d*, dashed gray trace) 3,5-dimethoxy-4-hydroxy cinnamic acid (sinapinic acid or SA). The vertical dotted line shows the emission wavelength of a nitrogen laser (337 nm). [Courtesy of Prof. Wendell P. Griffith (University of Toledo Department of Chemistry).]

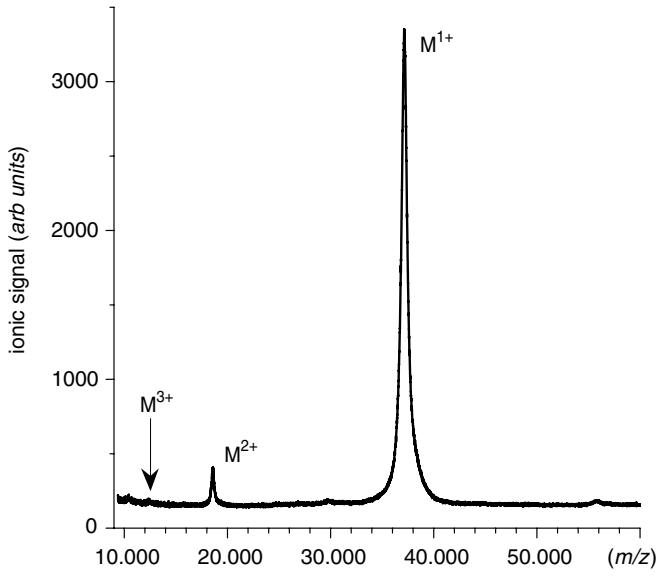


Figure 3.9. A conventional (UV) MALDI Time of flight (TOF) mass spectrum of human serum transferrin N-lobe. Compare the extent of multiple charging with that displayed in the ESI mass spectrum of the same protein (Fig. 3.6). [Courtesy of Rachael Leverence (currently at University of Wisconsin).]

with limited mass range can still be used for detection of smaller peptides produced by MALDI. Generally, MALDI surpasses conventional ESI in terms of sensitivity (detection levels in the low attomole range have been reported) and is more tolerant to salts. Superior sensitivity, relative simplicity of operation, and ease of automation have made it a top choice as an analytical technique for a variety of proteomics-related (high throughput) applications. On the other hand, MALDI mass spectra generally are not as reproducible as those obtained with ESI. Sample preparation is clearly the major critical factor in obtaining usable and reproducible data. This depends strongly on a number of components, including analyte type, choice of matrix, solvent, and added salts.

Like ESI, MALDI is a "soft ionization" method that enables intact macromolecular ions to be transported into the gas phase for analysis by MS. However, increased laser fluence often results in deposition of excessive energy, leading to analyte ion fragmentation. The extent of fragmentation often can be controlled by modulating the power of the laser beam, allowing this phenomenon to be used analytically as a means of producing structurally diagnostic fragment ions. It appears that collisional cooling in the plume region is the major suppressor of the metastable ion dissociation. Collisional cooling can be greatly enhanced by elevating the background pressure in the MALDI source

region to intermediate (~1 Torr) or high (1 atm) levels. Matrix-assisted laser desorption ionization at atmospheric pressure [AP-MALDI (65)] offers an additional advantage of allowing liquid matrices to be used (66). Importantly, AP–IR–MALDI can be used to analyze peptides using *liquid water* as a matrix (67,68), breaking the ESI monopoly in the field of biopolymer analysis directly in aqueous solutions. However, it still remains to be seen if this feature of AP MALDI will have significant impact in biophysics and structural biology.

3.3. MASS ANALYSIS

3.3.1. General Considerations: *m/z* Range and Mass Discrimination, Mass Resolution, Duty Cycle, and Data Acquisition Rate

Once the macromolecular ions have been produced and transferred to the vacuum, they can be selectively manipulated and detected using the vast arsenal of MS techniques. In this respect, macromolecular ions can be dealt with using the principles developed originally for handling small organic and inorganic ions. An important difference, however, is the large size of biopolymers, which makes certain aspects of ion manipulation and detection much more challenging.

Generally, several issues have to be considered to ensure high-quality mass measurement for macromolecular ions produced by ESI or MALDI. Perhaps the most important one is the requisite *m/z* range of a mass analyzer. For example, macromolecular ions generated by MALDI typically have few charges, hence the requirement that the mass analysis be carried out using an analyzer with an extended *m/z* range. This can be best accomplished by TOF mass analyzers, which will be briefly reviewed at the end of this chapter. Electrospray ionization, on the other hand, produces polymer ions with a significantly higher number of charges, as compared to MALDI. As can be seen from Figures 3.6 and 3.9, the most abundant ionic species of a 37 kDa protein *h*Tf/2N in the ESI mass spectrum carries 22 charges, while the most abundant ionic species of the same protein in the MALDI spectrum carries only one positive charge. This, of course, relaxes the requirements vis-à-vis the required *m/z* range of the analyzer employed for the detection of ESI generated biopolymer ions. In many cases, relatively inexpensive analyzers with a modest *m/z* range not exceeding 3000 would suffice (e.g., low-end quadrupole filters and ion traps). An important exception, though, is the situation when ESI MS is used to detect proteins under near-native conditions in solution. As we will see in Chapter 4, the number of charges accumulated by such protein ions is usually relatively low. As a result, *m/z* values of such "low charge density" ions will be very high, warranting the use of analyzers with an extended *m/z* range. In the extreme cases of large macromolecular complexes, an *m/z* range of up to 20,000 may be required.

The second important characteristic of the mass analysis process is the resolution (or mass resolving power) attained. Mass resolution defines the minimal difference between the *m/z* values of two ionic species that still allows a clear distinction to be made during the mass analysis. Historically, the two peaks of equal intensity were considered to be resolved if the valley between them was 10% of the maximum intensity. Hence, the classical definition of the resolution was

$$R = \frac{M}{\Delta M} \qquad (3\text{-}3\text{-}1)$$

where M is the *m/z* value of a particular ion peak, and ΔM is the width of this peak at 5% of its height. More recently, a much more forgiving definition of mass resolution was adopted, one using ΔM at 50% of the peak height. Regardless of the definition used, resolution is a function of *m/z*, rather than a fixed characteristic of a given instrument.

The observation of distinct isotopic species of peptides and proteins of progressively increasing size can be achieved by increasing the mass resolving power of the analyzer (Fig. 3.10). In some cases, ultrahigh resolution may allow the isobaric species to be separated and distinctly detected (14). Resolution is very important for accurate mass measurements, as in many cases it allows the monoisotopic mass to be accurately measured. However, the very low abundance of monoisotopic peaks of large proteins makes their detection all but impossible. In such cases, high-resolution mass measurements do not offer any significant advantage over "conventional" MS with modest resolution, unless the abundance of the monoisotopic peak is increased [e.g., by using isotope depletion during protein expression (69)].

Finally, issues related to the duty cycle of a mass analyzer and its data acquisition rate need to be taken into consideration when selecting the mass analyzer appropriate for a specific task at hand. Matrix-assisted laser desorption ionization, by definition, is a pulsed process and is best interfaced with "fast cycling" analyzers that can be synchronized with a laser (TOF, ion trap, etc.). Interfacing MALDI with "scanning" devices, (e.g., a magnetic sector MS), results in significant losses in sensitivity and requires extended acquisition times (70). On the other hand, ESI is a "continuous" ionization process, hence there is a wide variety of mass analyzers to which it can be interfaced. Still, the accumulation of ESI generated ions in an external "storage" device followed by a pulsed introduction into a TOF or ion trap mass analyzer greatly increases the duty cycle (and, therefore, sensitivity) of the analysis compared to a conventional scheme when the ESI source is interfaced directly to a slow scanning analyzer.

A very important advantage offered by MS as an analytical technique is the possibility to implement various

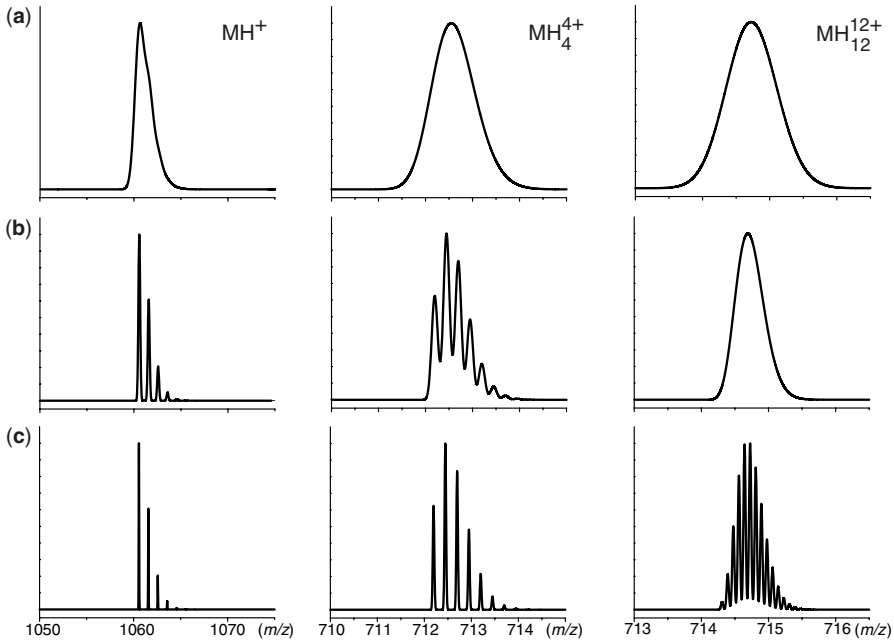


Figure 3.10. Simulated isotopic patterns three peptide ions: bradykinin (*a*), melittin (*b*), and ubiquitin (*c*) at a resolution of 1,000 (left), 5,000 (center), and 20,000 (right).

"hyphenated" techniques, ranging from direct (online) coupling of MS with separation techniques for increased selectivity and sensitivity of the analysis and implementation of tandem MS (MS/MS and MS^n) strategies for structural analysis.

3.3.2. Mass Spectrometry Combined with Separation Methods

The combination of separation techniques with mass spectrometric detection is a mature field, which has been reviewed in several excellent papers (71–73). In this section, we will only recount certain features of this combined technique that are particularly relevant to the biophysical experiments discussed in the following chapters. Electrospray ionization is a particularly attractive interface between liquid chromatography (LC) and MS due to the "continuous" nature of the ionization process, which allows analysis of the eluate content to be carried out in real time ("online" LC–MS). High-performance liquid chromatography–ESI MS is now a mature technology, with most commercial instruments available as integrated LC–MS systems. Current development of this technology follows two major routes: reduction of the scale and of the analysis time (71).

Miniaturization of the separation step allows the amount of sample needed for analysis to be reduced very dramatically. Even for conventional ESI sources the typical flow rate is usually on the order of several μL/min. Such flow rate requirements are optimally matched by micro-HPLC systems, while HPLC systems with higher flow rates require either postcolumn eluate flow split prior to introduction to the ESI source or utilization of "ionspray" sources capable of handling high eluate flow rates. Besides matching the flow rates of the HPLC and ESI MS, one needs to be aware of another important technical issue related to solvent compatibility. Reversed-phase LC separation of peptides and proteins is commonly performed using trifluoroacetic acid (TFA) as an eluent modifier. This acid is a very strong electrolyte ($pK_a < 0.5$), and hence acts as an ion-pairing agent providing superior chromatographic characteristics (separation and peak shape), but its presence in the solvent usually decreases the quality of ESI MS data. If necessary, the detrimental effect of TFA on ESI MS measurements can be reduced or eliminated using two different approaches. In the first approach, TFA concentration in the eluate is reduced by mixing it with a flow of liquid containing volatile ion-pairing agents, a step that can be carried out either "postcolumn" or in the ESI source. More sophisticated schemes utilize a second chromatographic step, whose mobile phase is modified with volatile reagents (74). An alternative approach utilizes weaker acids (typically formic or acetic acids) instead of TFA during the separation step. An unavoidable loss of chromatographic fidelity is compensated by MS detection, which allows chromatographically poorly resolved analytes to be easily distinguished, based on the differences in their masses.

Traditionally, LC was not viewed as a "fast" method of chemical analysis and was poorly suited for high-throughput applications. The time-consuming nature of chromatography has stimulated design of methods that improve the speed of separations (75). As we will see in upcoming chapters, fast HPLC separation of peptic fragments is crucial for high-quality analysis of protein dynamics by monitoring hydrogen–deuterium exchange (HDX) reactions in a site-specific fashion. Significant improvements in the speed of separation can be achieved by using smaller and nonporous or superficially porous particles as a column packing material. The potential for fast separations can be further increased by using capillary columns, monolithic columns, open tubular columns, and small diameter packed capillaries (75). Elevated temperatures can also be effective for decreasing analysis times. However, the HDX HPLC–MS measurements (the most demanding fast-LC applications in biophysical studies so far) are usually carried out at low temperatures to avoid excessive exchange of deuterium atoms between the peptides and the mobile phase during the separation step. Increasing the separation speed is usually achieved at the expense of separation quality. Again, this is usually compensated by the high detection specificity of MS. One recently introduced technology that holds particular promise for biophysical studies is ultraperformance liquid chromatography (UPLC), which utilizes columns that can be operated at much higher pressure compared to HPLC. This has been shown to offer significant improvements in sensitivity, speed, and resolution (76). A great deal of useful practical information on various LC–MS techniques can also be found in recently published books (77,78).

Capillary electrophoresis (CE) is another separation technique that can be interfaced directly with ESI MS. However, the design of the CE–ESI MS interfaces is technically more difficult due to the problems associated with the high voltage utilized by CE for analyte separation (71,79,80). Although the number of applications of online CE–ESI MS systems is still lagging behind that of LC–MS, the former technique is steadily gaining popularity due to its superior sensitivity and resolution. An interested reader is referred to an excellent recent review on this subject (81). Interfacing of ESI MS with other types of separations, such as immobilized metal ion affinity chromatography [IMAC (82)], ion exchange, or bioaffinity chromatography (83) is also possible, although the coupling cannot be direct and involves an intermediary HPLC step. A concise overview of these multidimensional chromatographic techniques can be found in recent reviews (84,85).

3.4. TANDEM MASS SPECTROMETRY

One of the most attractive features of both ESI and MALDI for biomolecular analysis is their ability to generate intact macromolecular ions in the gas phase. While molecular weight information is very useful, it is not sufficient in most instances for unequivocal identification of even small peptides. Furthermore, while high-resolution mass measurements can sometimes reveal the molecular formula of a small peptide, they do not provide any information on its covalent structure. The latter can be obtained by inducing dissociation of the molecular ion and measuring the masses of the resulting fragment ions. Since most proteins and peptides are linear polymers, cleavage of a single covalent bond along the backbone generates a fragment ion (or two complementary fragment ions) classified as an *a*-, *b*-, *c*- or *x*-, *y*-, *z*-type (86,87), depending on (1) the type of bond cleaved and (2) whether the fragment ion contains an N- or C-terminal portion of the peptide (Fig. 3.11). Cyclic and disulfide-linked polypeptides are a special case, since a single bond cleavage does not necessarily produce distinct (physically separated) fragments. Nomenclature for cyclic peptide fragmentation can be found in (88).

3.4.1. Basic Principles of Tandem Mass Spectrometry

Ion dissociation can often be carried out in the ionization source, for example, by increasing the desolvation potential in the ESI interface ["nozzle-skimmer dissociation", first observed by Alexandrov et al. (38)] or by using high laser power in MALDI. If the sample is homogeneous, such in-source fragmentation can be used for biopolymer sequencing (Figs 3.12 and 3.13). In most cases, however, the sample to be analyzed is a rather heterogeneous mixture, and the "in-source" fragmentation spectrum would be very difficult to interpret due to the presence of fragment ions derived from different precursor ions. One way to resolve this problem is to use an online separation of analytes prior to their introduction to the ionization source (e.g., an HPLC – ESI MS interface discussed in Section 3.3). However, a much more flexible and powerful solution to this problem utilizes the mass spectrometer itself as an ion separation device, which allows the fragmentation of mass-selected precursor ions to be carried out in a variety of ways. This approach, known as tandem MS, or MS/MS, is now a commonly accepted way to obtain information on covalent structure of a range of biopolymers, but is particularly well suited for analysis of polypeptides. Tandem mass spectra of peptide ions are easily interpretable, and even though the process of mass-selection greatly reduces the overall ionic signal intensity, the signal-to-noise ratio (S/N) achieved in MS/MS experiments is often better than in the MS1 experiment due to the elimination of chemical noise (Fig. 3.14). As used above, the term "tandem" implies physical separation of the precursor ion prior to fragmentation (MS1 selects the precursor ion for consecutive fragmentation and MS2 records a spectrum of

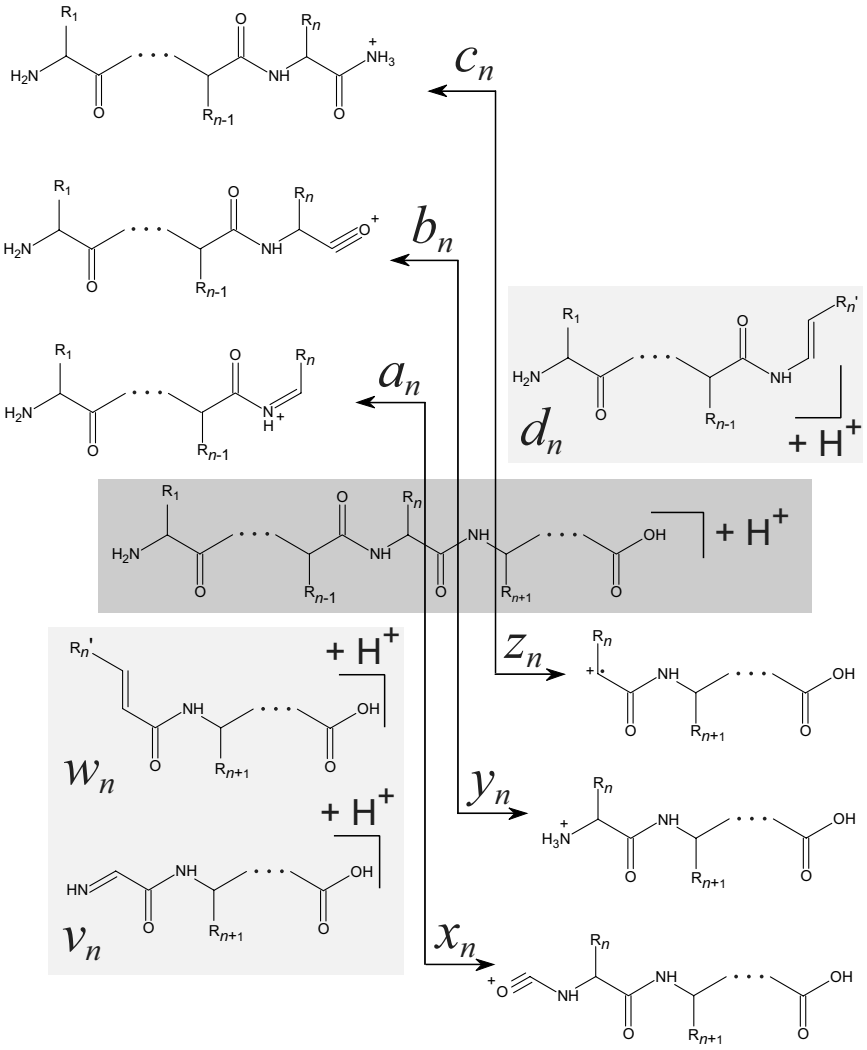


Figure 3.11. Biemann's nomenclature of peptide ion fragments (86). Fragment ions shown in gray boxes correspond to either complete or partial loss of the side chains and are usually observed only in high-energy collision-induced dissociation (CID).

the resulting fragment ions); nonetheless, the two stages (MS1 and MS2) can be combined in one step. As we will see in the following sections, MS/MS experiments can be carried out without physical separation of the chosen precursor ion from other ions generated in the source. A generalized definition of MS/MS extends to any experiment that defines an *m/z* relationship between a precursor and a fragment ion (89).

3.4.2. Collision-Induced Dissociation: Collision Energy, Ion Activation Rate, and Dissociation of Large Biomolecular Ions

The majority of MS/MS experiments employ various means of increasing internal energy of the precursor ions to induce the dissociation of covalent bonds in the gas phase. Collisional activation (conversion of a fraction of the ion kinetic energy to vibrational excitation upon its collision with a neutral molecule) is perhaps the most widely used method of elevating ion internal energy. This technique is usually referred to as CID or collision-activated dissociation (CAD) (90). Two distinct regimes of collisional activation are usually recognized: high and low energy. The low-energy collisional regime refers to a broad range of ion kinetic energy prior to collision (usually in the sub-electron-volt range) and typically requires multiple collisions in order to accumulate enough internal energy to afford dissociation of a covalent bond. Therefore, low-energy CID is a slow process, with the activation time typically exceeding 10 ms (91). Note that *low-energy* and *slow-heating fragmentation* are not necessarily synonymous; a good

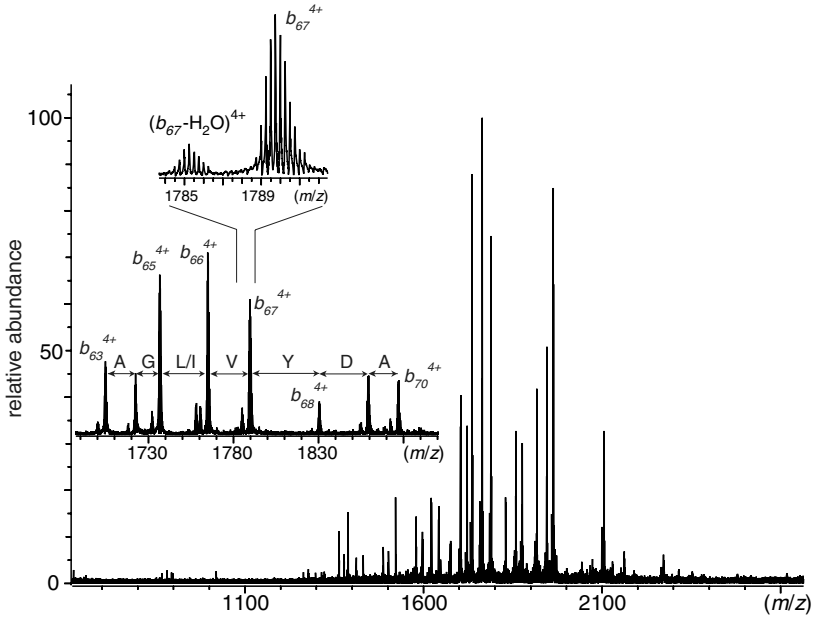


Figure 3.12. Prompt ion fragmentation in ESI MS: mass spectrum of *h*Tf/2N acquired under elevated skimmer potential. A mass spectrum of this protein acquired under gentle conditions in the ESI interface is shown in Figure 3.6.

example is the so-called SORI CID (sustained off-resonance irradiation CID, which will be discussed in Section 3.5.5). While the energy of each collision in SORI may be relatively high, the frequency of such collisions is typically very low. As a result, internal energy accumulation is slow due to the efficient radiative cooling of the ions between consecutive collisions (92).

Low-energy CID spectra of peptide ions are usually dominated by fragment ion peaks corresponding to *b*- and *y*-ions (Fig. 3.15*a*,*b*). One of the potential consequences of multiple collisions is the formation of internal fragment ions, whose presence in the CID spectra makes their interpretation more difficult, and the greater propensity for rearrangement prior to fragmentation. However, in many cases it is possible to avoid these processes altogether by selecting appropriate experimental conditions.

High-energy CID is usually carried out by accelerating ions to several kiloelectronvolts prior to colliding them with

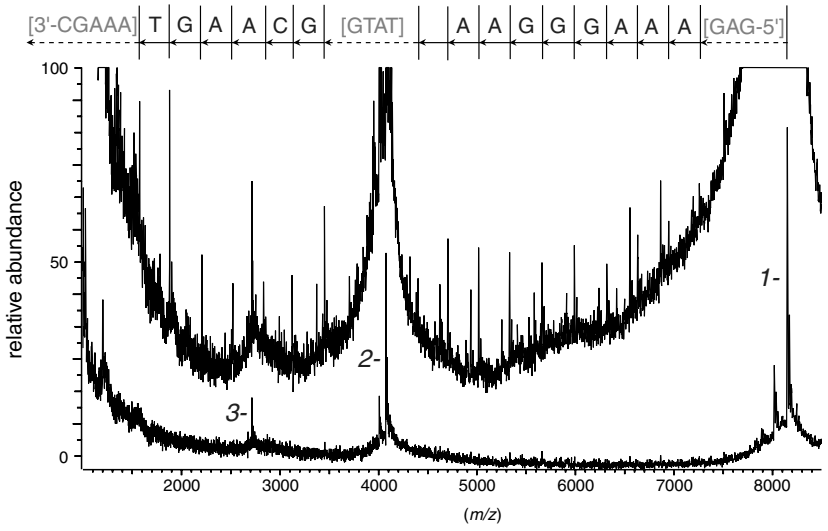


Figure 3.13. Prompt fragmentation in MALDI MS: UV MALDI spectra of an oligonucleotide strand acquired at increased (top trace) and moderate (bottom trace) laser power.

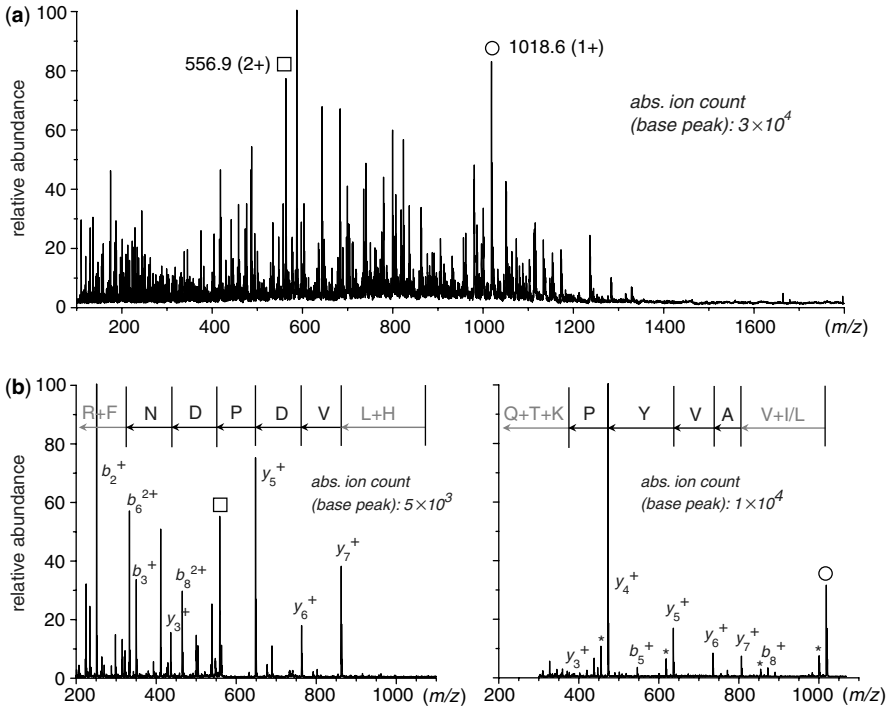


Figure 3.14. Tandem mass spectrometry (CID) of tryptic peptides from a digest of *Gadus morhua* hemoglobin. Two peptide ions marked with a square (doubly charged β_{97-105}, LHVDPDNFR, *m/z* 557) and a circle (singly charged α_{33-41}, LVAVYPQTK, *m/z* 1019) in a spectrum of the mixture (*a*) were mass selected and fragmented in a quadrupole trap. In each case, a series of abundant *y*-ions provides almost complete sequence coverage of a peptide ion (*b*). [Courtesy of Prof. Wendell P. Griffith (University of Toledo Department of Chemistry).]

neutral targets (typically atoms of inert gases). For smaller ions, a single high-energy collision is often sufficient to cause dissociation of a covalent bond. However, as ion size increases, so does the number of vibrational degrees of freedom among which the excitation is distributed. Furthermore, collisional energy in the center-of-mass frame decreases with increasing mass of the projectile ion when the kinetic energy of the projectile and the mass of the neutral target are fixed. Nonetheless, in many cases it is possible to obtain nearly complete sequence coverage even for relatively large (up to 5 kDa) polypeptide ions (93,94). In sharp contrast to the fragmentation patterns observed with multiple low-energy collisions, fragment ions produced by high-energy CID are amply formed by extensive cleavages of the backbone not limited to the amide bonds, as well as the side chains (Fig. 3.15*c*). Dissociation of the covalent bonds occurring along the amino acid side chains leading to elimination of either the entire side chain (*v*-ions) or a portion of it (*d*- and *w*-ions) is unique to high-energy CID and can provide information on the identity of isomeric side chains, for example, by distinguishing between leucine and isoleucine (95).

Finally, note that the vast majority of CID work pertaining to bioanalysis deals with cationic species. Collisional activation of negative polypeptide ions frequently leads to loss of side chains, leaving the backbone intact (96). Although this feature does not allow direct application of negative-ion CID to protein sequencing and, therefore, greatly diminishes its diagnostic value, there is one important exception when it can and does provide useful structural information. Cysteine is one of the amino acids whose side chain is prone to cleavage by CID in the negative-ion mode (97), which allows this technique to be used for cleaving disulfide bonds in the gas phase, a task that cannot be accomplished by conventional CID (98). Importantly, cleavage of disulfides in negative-ion CID also produces a very clear "signature triad" in the fragment-ion spectrum (Fig. 3.16), which allows the cystine-containing products to be easily identified. This unique feature makes negative-ion CID a very useful tool in analyzing disulfide patterns in proteins (99–101), which will be considered in more detail in Chapter 4.

3.4.3. Surface- and Photoradiation-Induced Dissociation

One of the major factors limiting the fragmentation yield of CID is the low efficiency of energy conversion (from translational to internal degrees of freedom). Collisional

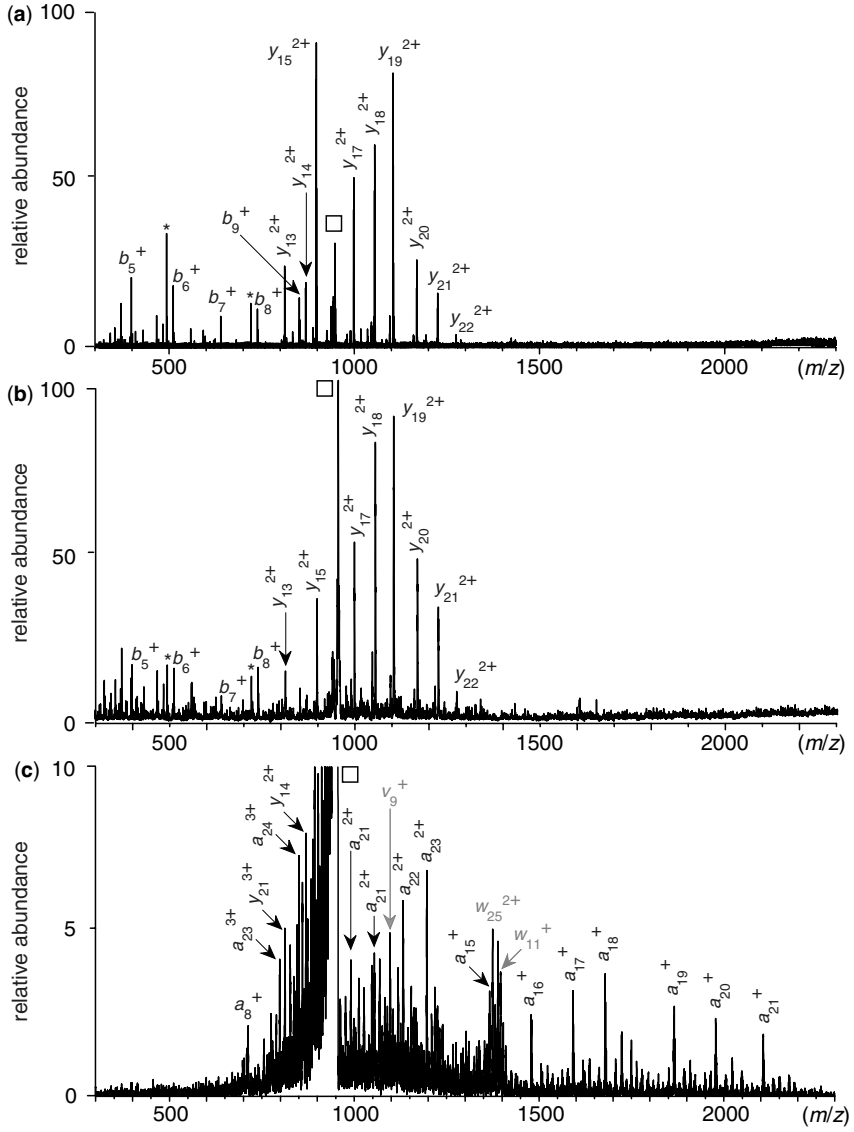


Figure 3.15. Low versus high-energy peptide ion fragmentation. Low-energy CID spectra of melittin (3+ charge state) were acquired with an FT ICR MS (SORI CID, *a*) and a quadrupole ion trap (*b*). High-energy CID spectrum was acquired with a magnetic sector MS (B/E linked scan mode, *c*). Only the most abundant fragment ions are labeled in the spectra. [Courtesy of Anirban Mohimen (currently at Vertex Pharmaceuticals) and Joshua Hoerner (currently at Merck Pharmaceuticals).]

activation of large biomolecular ions is particularly problematic due to the large disparity between the masses of the target (neutral atom or molecule of collision gas) and the projectile (ion), with the mass of even the heaviest target (Xe atoms) being at least an order of magnitude below that of a relative modest biomolecular ion. Such mass disparity results in relatively modest collisional energy in the center-of-mass frame even if the energy of the projectile in the laboratory frame is very high. One way to circumvent this problem is to employ a collision target with large (ideally, infinite) mass. Practical implementation of this approach led to the development of surface-induced dissociation (SID), a fragmentation technique that utilizes ion-surface collisions as a means to convert ion kinetic energy to internal excitation (102,103). However, the fraction of energy deposited into the ion upon such collision very seldom approaches unity, as it is largely determined by the mass of the chemical moiety representing an immediate collision partner for the ion impacting the surface (104). One practical aspect of SID that makes its applications rather limited is the difficulty of refocusing the fragment ions after collision with the surface.

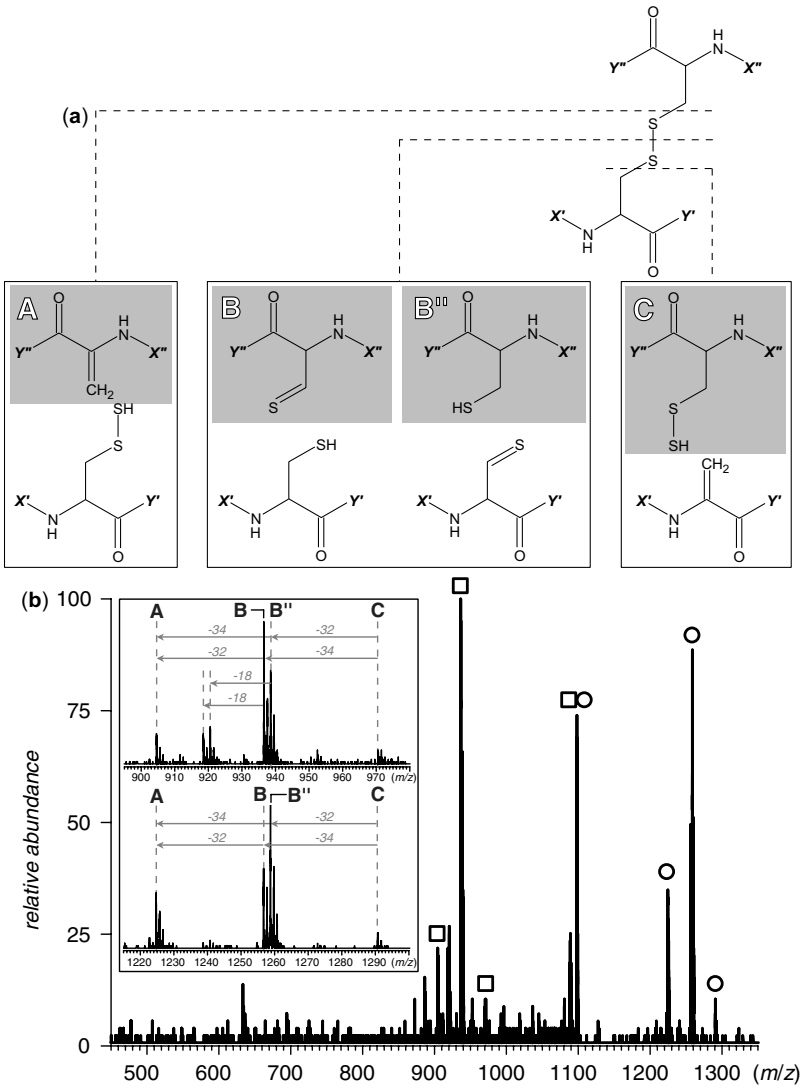


Figure 3.16. (*a*) Nomenclature for fragment ions produced by cleavage of disulfide bonds in peptide dimer anions in the gas phase. (*b*) Mass spectrum of fragment anions generated by CAD of a disulfide-linked peptide dimer [8–18]–[43–50] derived from digestion of *h*Tf/2N with trypsin. The spectrum was obtained with a hybrid QqTOF mass spectrometer (CAD carried out in an rf-only quadrupole). Open squares and circles indicate ions corresponding to intact peptide monomers produced by dissociation of the disulfide bond in the peptide dimer ion without any backbone cleavages (zoomed views are shown in insets). Adapted from (100).

Ion-neutral collision is not the only process that can be used to increase ion internal energy. In principle, any exothermic process can be employed for ion activation purposes. For example, photoexcitation of ions in the gas phase often leads to their dissociation. Dissociation of large macromolecular ions is most effective when infrared photons are used for the excitation [a technique known as *IR multiphoton dissociation* (IRMPD (105)]. Unlike IRMPD, photodissociation induced by UV light has not enjoyed much popularity as a tool to study biomolecular structure despite the numerous applications focusing on small ions. The situation began to change several years ago following a realization that a close match between the energy of the photon utilized for ion activation and the strength of certain chemical bonds may lead to highly selective excitation and dissociation processes. For example, the wavelength of maximum absorption of a disulfide bond is close to 150 nm, which allows very selective cleavage of thiol–thiol linkages in disulfide-bound peptide dimer ions to be achieved upon their irradiation with a 157 nm laser beam (106). In addition, UV induced dissociation of peptide ions at this wavelength may also produce cleavage

of a C(α)−C(O) bond in the backbone, leading to formation of *a* - and *x*-ions, as well as loss of side chains resulting in *v*- and *w*-type fragments (107). Another UV band showing significant promise vis-à-vis fragmentation of peptide ions is 193 nm, which has been shown to produce facile cleavage of the backbone (leading to formation of *a*-, *b*-, *c*- *x*-, *y*-, and *z*-ions), as well as side-chain loss (*v*- and *w*-fragments) (108).

Although UV-induced photodissociation offers multiple advantages over collisional activation and IRMPD (e.g., by offering high-energy and very fast energy deposition with a potential to provide highly selective chemical bond excitation and cleavage) (109), there are also certain drawbacks. These include relatively high cost of instrumentation, safety concerns, and apparent variations in fragmentation patterns observed by different groups (109,110). Nevertheless, it seems very likely that this technique will become an indispensable part of biological MS (including biophysical studies) in the very near future, a development that would certainly be catalyzed by its anticipated commercialization (109).

3.4.4. Electron-Based Ion Fragmentation Techniques: Electron Capture Dissociation and Electron Transfer Dissociation

Gas-phase ion fragmentation processes involving electron transfer form the foundation of two other recently introduced techniques that are closely related to each other, namely, *electron capture dissociation* (ECD) (111) and *electron transfer dissociation* (ETD) (112). Fragmentation induced by ECD and ETD is fast and proceeds in many cases with little or no energy partitioning. From the analytical standpoint, it produces a tremendous advantage of preserving labile groups (e.g., carbohydrate chains and other post-translational modifications (PTM) that usually do not survive collisional activation), while cleaving the peptide-ion backbone, where it mostly generates *c*- and *z*-fragments. In fact, the backbone cleavage in many cases occurs without causing dissociation of noncovalent complexes (113), a feature that will be explored in depth in later chapters. Very often ECD and ETD provide more extensive sequence coverage in polypeptides compared to conventional CID (Fig. 3.17). Another very attractive feature of electron-based fragmentation techniques is their ability to cleave disulfide bonds in the gas phase (114), a challenging task when other methods of ion activation are employed, at least in the positive ion mode. Since the fragmentation patterns produced by ECD and ETD are frequently complementary to the CID generated fragments (115), it is beneficial to utilize them in multistage fragmentation (the so-called MS^n experiments), where the initial fragmentation event is caused by collisional activation, followed by subjecting the resulting CID fragments to ECD or ETD. The combination of CID with either ECD or ETD is now offered in several commercial instruments.

3.4.5. Ion-Molecule Reactions in the Gas Phase: Internal Rearrangement and Charge Transfer

Ion activation times in the gas phase can range from $<10^{-15}$ to $>10^{3}$ s, depending on the particular activation technique used to induce fragmentation, physical size of the ion

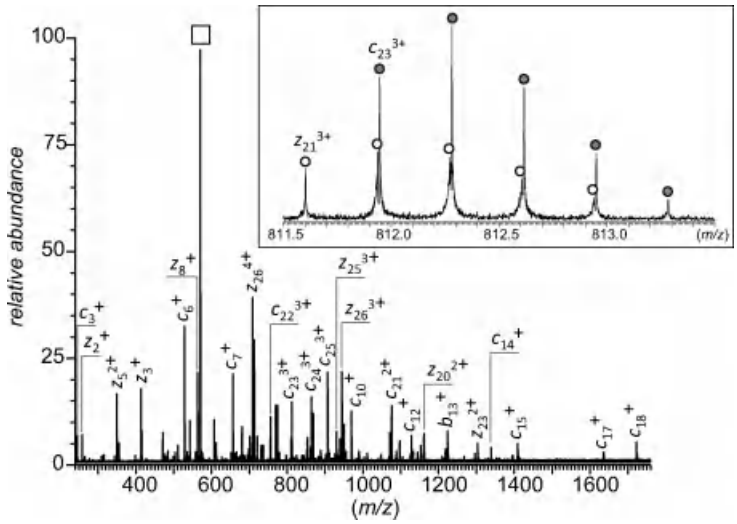


Figure 3.17. The ECD mass spectrum of melittin (charge state 5+ labeled with a square was selected as a precursor ion) obtained with a 7T FT ICR mass spectrometer. Separation of isobaric c_{23} and $z_{23}+1$ fragment ions (inset) is possible due to the high resolving power of the mass analyzer. [Data courtesy of Dr. Rinat Abzalimov (University of Massachusetts-Amherst). CAD mass spectra of melittin are shown in Figure 3.15.]

(number of internal degrees of freedom), as well as the time frame of the experiment (91). In the case of a slow activation process, a covalent bond cleavage may be preceded by an internal rearrangement. One particular type of such rearrangement is hydrogen scrambling, which will be discussed in some detail in Chapter 5.

Charge transfer is another gas-phase process that frequently occurs in the ESI interface region. Although charge transfer reactions do not usually result in ion fragmentation, they obviously affect the appearance of charge state distributions in ESI mass spectra. As we will see in Chapter 4, charge-state distributions of protein ions are often used to assess "compactness" of protein structures in solution. Therefore, close attention needs to be paid to gas-phase processes in order to avoid misinterpretation of the ESI MS data. In some instances, it is possible to trap ions of different polarities simultaneously and confine them to a small volume (116). Electrostatic attraction between the ions of opposite charges may lead to a variety of ion–ion reactions, some of which will be considered in Chapter 8.

3.5. BRIEF OVERVIEW OF COMMON MASS ANALYZERS

The mass analyzer is the part of a mass spectrometer where the ions are separated according to their *m/z* values. As outlined in Section 3.4, combination of two (or even more) mass analyzers often allows spontaneous or induced fragmentation of mass-selected ions to be studied using methods of MS/MS. Certain types of mass analyzer allow tandem experiments to be carried out without utilization of a second analyzer ("tandem-in-time" as opposed to "tandem-in-space" MS). There is a wide range of mass analyzers differing in their compatibility with various ion sources, ability to handle ions of certain types, analytical figures of merit, user-friendliness, and, of course, price tag. This section attempts to provide a brief review of mass analyzers that are most popular in the bio-MS community. A brief discussion of each analyzer will include principles of its operation, compatibility with ESI and MALDI sources, *m/z* limitations, and commonly attainable resolution, and tandem capabilities. A much more comprehensive discussion of mass analyzers, which also includes devices and designs not covered in this section, can be found in an excellent review by McLuckey and Wells (89). A more in-depth discussion of physical aspects of mass analysis of ions can be found in other review articles (117,118).

3.5.1. Mass Analyzer As an Ion Dispersion Device: Magnetic Sector Mass Spectrometry

The idea to use a combination of an electrostatic and a magnetic field as a means of ion separation in space (dispersion) was introduced by J.J. Thomson in his parabola mass spectrograph (1). A simpler and more efficient method of ion separation in homogeneous magnetic fields was introduced several years later by Dempster (119). This instrument became a prototype of the highly successful *magnetic sector* mass analyzer, which is widely used in MS to this day. According to Eq. 3-1-1, an ion introduced to the magnetic field orthogonally to the field will follow a circular trajectory whose radius will be determined by the ion's *m/z* ratio, its velocity *v* and the magnetic field strength *B*. In other words, the magnet will act as a momentum separator (or, more precisely, a momentum-to-charge (*mv/z*) separator). If all ions of the same charge have been accelerated to the same kinetic energy (KE) prior to their introduction into the magnetic field (i.e., $KE = {}^1/_2\, mv^2 = zeV$), then the radius of circular trajectory for each ion will be uniquely defined by its *m/z*.

Later modifications of the magnetic sector mass analyzers have led to significant improvements in resolution and other performance characteristics. Perhaps the most important among such modifications was the implementation of a "double-focusing" scheme(120), which adds another analyzer (a sector with radial electrostatic field termed Electrostatic Analyzer, ESA) acting as a "kinetic energy separator". A mass spectrum is usually obtained by scanning the magnetic field strength over a desired range, while the electrostatic field of the ESA remains constant (linked to the acceleration voltage V_0 to allow passage of ions whose kinetic energy-to-charge ratio is equal to eV_0). In addition to a dramatic increase in mass resolution, the presence of a second analyzer allows MS/MS experiments to be carried out. For example, a fragment ion m_f^{zf+} produced upon dissociation of a metastable ion $m_0{}^{z0+}$ immediately following their acceleration (in the so-called first field-free region, 1FFR in Fig. 3.18) will have the same velocity as the precursor ion itself; however, the momenta and the kinetic energies of these two ions will be different. Here, both the momentum and the kinetic energy of the fragment ion relate to those characteristics of the precursor ion as m_f/m_0. Therefore, in order to guarantee the passage of the fragment ion through *both* magnetic sector and ESA, both fields have to be reduced by a factor of m_f/m_0 compared to those needed for the passage of the precursor ion. If both fields are scanned, while their ratio remains constant, a full range of fragment ions originating from the same precursor will be detected, an experiment commonly known as a *B/E scan* or *linked scan*). The representation of the *B/E* scan line on the (*B, E*) plane connects the (B_0, E_0) point with the coordinate origin, where B_0 is the magnetic field strength needed for passage of the intact precursor ion through the magnetic sector and E_0 is the electrostatic field strength required for passage of any "fully accelerated" ion (KE/z = eV) through ESA (Fig. 3.19). Obviously, if $z_0 > 1$, the *B/E* scan should start at the point (z_0B_0, z_0E_0), otherwise none of the

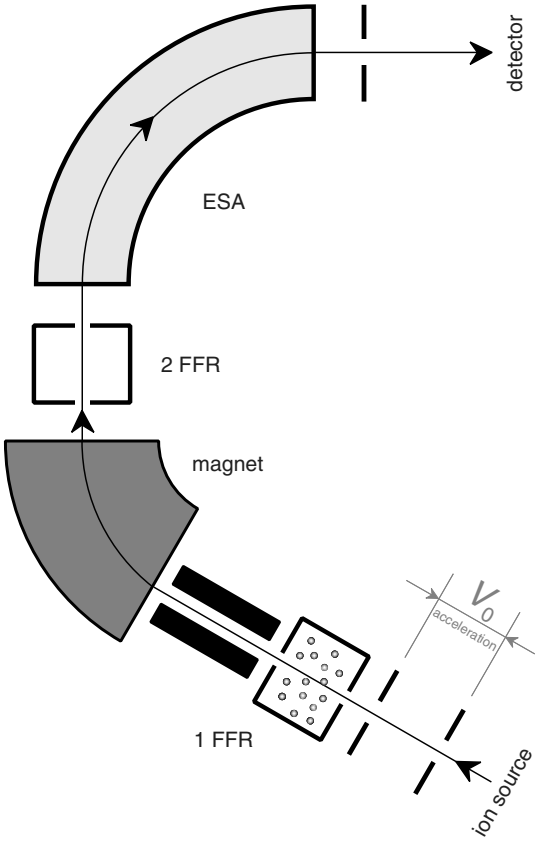


Figure 3.18. Schematic representation of a magnetic sector MS (BE geometry). Collision gas in the first field-free (FFR) region (as shown in the diagram) is used only for acquisition of CID spectra (B/E, B^2/E and CNL scans).

fragment ions whose charge is less than that of the precursor ion would be detected. This is exactly how the fragmentation spectrum of melittin (Fig. 3.15c) was acquired.

The *B/E* scan is an example of a "pseudo-tandem" MS experiment (the precursor ion is not mass-selected prior to its dissociation). Other examples of pseudo-tandem MS are the *B^2/E scan*, which provides a means to detect all ions giving rise to a certain fragment (*precursor ion scan*), and the *constant neutral loss scan* (CNL), which measures all fragments related to parent ions by loss of the same mass. Physical selection of the precursor ion prior to its dissociation can be achieved by placing a collisional cell after the magnetic sector (second field-free region, Fig. 3.18). Since the kinetic energy of each fragment ion will be a fraction (m_f/m_0) of the precursor kinetic energy, scanning the electrostatic field of the ESA will allow a mass spectrum of all fragment ions to be acquired. Although this technique (termed MIKES) has rather limited analytical use (due to very low-mass resolution), it can provide information on macromolecular-ion geometry. A combination of at least three sectors (e.g., *BEB*) is required in order to obtain high-resolution MS/MS data.

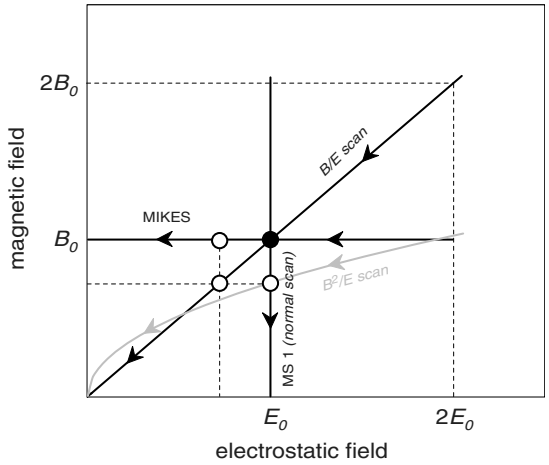


Figure 3.19. Graphic representation of various scans used in sector mass spectrometers similar to one whose schematics is shown in Figure 3.18. The filled circle indicates the position of a precursor ion and the open circles indicate positions of one particular fragment ion as detected in the course of various experiments: magnetic field scan (prompt fragment in MS1), as well as a linked (*B/E*) scan and an electrostatic field scan mass-analyzed ion kinetic energy spectroscopy, (MIKES). The gray line represents the precursor ion scan *(B^2/E)*.

The magnetic sector MS is a very flexible analytical device that allows a variety of experiments to be carried out without any hardware modification. Sectors allow high-resolution measurements to be carried out over a wide *m/z* range. Ion fragmentation experiments are carried out in a high collisional energy regime, enabling observation of fragments that are not usually detected when most other analyzers are used (e.g., *d*-, *v*-, and *w*-ions in peptide fragmentation spectra). However, magnetic sector instruments have significant disadvantages as well. Since the data acquisition rate is usually limited by magnet scan rate, sector MS cannot be interfaced readily with pulsed ionization sources (e.g., MALDI). Even when interfaced with a continuous ionization source (e.g., ESI), the sensitivity of the analysis is often limited by the unfavorable duty cycle (particularly when the data acquisition is carried out over a wide *m/z* range). Nevertheless, it is possible to use these instruments for very demanding applications requiring access to a high *m/z* range (121).

3.5.2. Temporal Ion Dispersion: Time-of-Flight Mass Spectrometer

The concept of TOF mass analyzers was first introduced over half a century ago under the names "time dispersion mass spectrometer" (122) and "ion velocitron" (123). The basic principle of TOF MS is very simple (Fig. 3.20): ions of different mass are accelerated to the same kinetic energy (by traversing a potential difference *V*) within a very short period of time and introduced into a field-free "drift region"

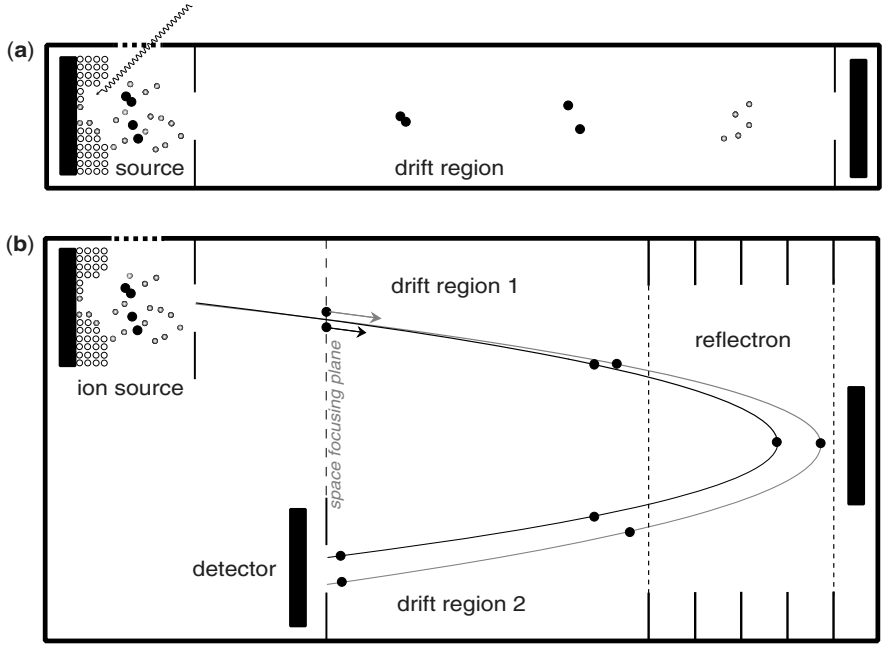


Figure 3.20. Schematic diagrams of linear (*a*) and single-stage reflectron (*b*) TOF mass spectrometers.

(or flight tube). If the initial velocities of all ions are neglected, then the final velocity of each ion in the drift region will be uniquely determined by its *m/z* and the acceleration potential U_0:

$$v = \sqrt{\frac{2zeU_0}{m}} \tag{3-5-1}$$

If the duration of the pulse during which ions were introduced into the drift region is very short, then ions arriving at the detector plane will be grouped according to their *m/z* values:

$$t = \frac{D}{v} = \sqrt{\frac{m}{2zeU_0}} \cdot D \tag{3-5-2}$$

where D is the length of the drift region and t is the time required to traverse this region. Therefore, measuring the ionic signal intensity as a function of the "arrival time" would allow the mass spectrum to be recorded. In reality, this simplistic approach would lead to very poor mass resolution, mostly due to significant spread in the initial kinetic energies of ions. A "correction" for the initial energy distribution can be done using an "ion mirror" or reflectron (124). The principle of reflectron operation is illustrated in Figure 3.20: If two ions have identical mass and charge, but different velocities, the faster ion will penetrate deeper into the decelerating region of the reflectron. As a result, its trajectory path will be longer. After its reemergence from the reflectron, this ion would still have a higher velocity, but it will be "lagging" behind the slower ion due to the extra time spent in the decelerating region. It is easy to show that the total "travel time" (from the source to the detector) of an ion having initial kinetic energy eV can be calculated as:

$$t = \sqrt{\frac{m}{2ze(V+U_0)}}\left(D_1+D_2+4d\right) \tag{3-5-3}$$

where D_1 and D_2 are the lengths of the "upstream" and "downstream" drift regions and d is the reflectron penetration depth (a function of eV). It is possible to adjust the reflectron parameters ($D_1 + D_2$ and U_0) in such a way so that the flight times become independent (within a narrow range) of the initial kinetic energy eV. The highest mass resolving power is attained when $D_1 + D_2 = 4d$, that is, the ion spends equal time in the reflectron and the field-free drift regions (125). Although such a "single-stage" reflectron can only perform first-order velocity focusing of ions initially located at the space focal plane (Fig. 3.20), second-order focusing can be achieved using a double-stage ion mirror. Theoretically, it is possible to achieve an ideal focusing (totally independent of the ion kinetic energy) by using a continuous quadratic field (without the field-free drift region). The quadratic ion mirror is impractical for a variety of reasons, however, its modification known as a "curved-field" reflectron (126) has become very popular. While a detailed discussion of ion focusing techniques in TOF MS is beyond the scope of this book, an interested reader is referred to an excellent book by Cotter (125), as well as several recent tutorials on the subject (127).

Tandem Experiments with TOF MS. Dissociation of metastable ions in the field-free region of a TOF mass analyzer (usually referred to as post-source decay, PSD) gives rise to fragment ions having the same velocity as their precursor. As a result, it would be impossible to distinguish such fragment ions from the intact precursor ions in a linear TOF mass spectrum. The situation is very different if a reflectron TOF MS is used for mass analysis. Although the velocity of the fragment ions produced in the first drift region would still be the same as that of the intact precursor ions in both field-free regions, they will be "turned around" in the decelerating field of the reflectron much faster. The resultant time of flight of the fragment ion m_f will be (125)

$$t = \sqrt{\frac{m}{2zeU_0}}\left(D_1+D_2+4\frac{m_f}{m}d\right) \qquad (3\text{-}5\text{-}4)$$

Mass selection of the precursor ions is usually accomplished in the first drift region by simple electrostatic gating. To achieve adequate and proper focusing of fragment ions of different masses, a series of spectra have to be collected at different decelerating potentials. Alternatively, a curved-field reflectron can be used to obtain a single PSD spectrum with product ions from a wide *m/z* range focused simultaneously (128).

Overall, fragmentation spectra provided by PSD are imperfect, since the activation of the ions takes place in the ion source and dissociation of the metastable ions can occur during acceleration, which has a detrimental effect on resolution. Furthermore, precursor ion selection cannot be accomplished with high precision, since it is carried out in a region where ions are generally out of focus (125). This problem can be circumvented by using tandem (TOF–TOF) mass spectrometers, where the drift regions are separated by a collision cell, in which fragmentation of the mass-selected ions is carried out using CID (129,130) or by other means of ion activation (131). Finally, interfacing TOF with other mass analyzers provides an opportunity to carry out tandem experiments with a hybrid mass spectrometer.

The advantages offered by TOF mass analyzers (ideal compatibility with "pulsed" ionization sources, duty cycle close to 100%, high ion transmission efficiency, extended *m/z* range, ability to carry out very fast analyses with repetition rates up to 500 kHz) make them extremely popular mass analyzers, which are uniquely suited for a variety of applications, many of which will be discussed elsewhere in this book.

3.5.3. Mass Analyzer As an Ion Filter

The ion separating properties of "mass filter" devices result from their ability to selectively maintain stable trajectories for ions of certain *m/z* ratios, while the others become unstable. The idea of using quadrupolar electrical fields as a means of "filtering" ions according to their *m/z* ratios was introduced and implemented in the 1950s (132–134). A quadrupole mass spectrometer acts as a "tunable" mass filter that transmits ions within a narrow *m/z* range (Fig. 3.21). A quadrupolar electrical field is usually configured using four parallel electrodes (metal rods of circular or, ideally, hyperbolic cross-section) that are connected diagonally. A periodic potential of the form

$$\Phi_0 = \pm(U-V\cos\omega t) \qquad (3\text{-}5\text{-}5)$$

is applied to each pair of electrodes, resulting in a periodic hyperbolic field configuration in the (x, y) plane:

$$\Phi(x,y) = (U-V\cos\omega t)\cdot\frac{x^2-y^2}{2r_0^2} \qquad (3\text{-}5\text{-}6)$$

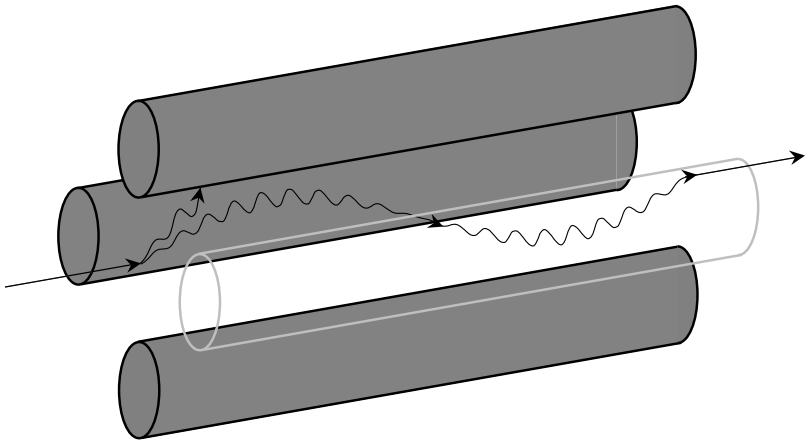


Figure 3.21. Schematic representation of a quadrupole mass filter with examples of stable and unstable ion trajectories.

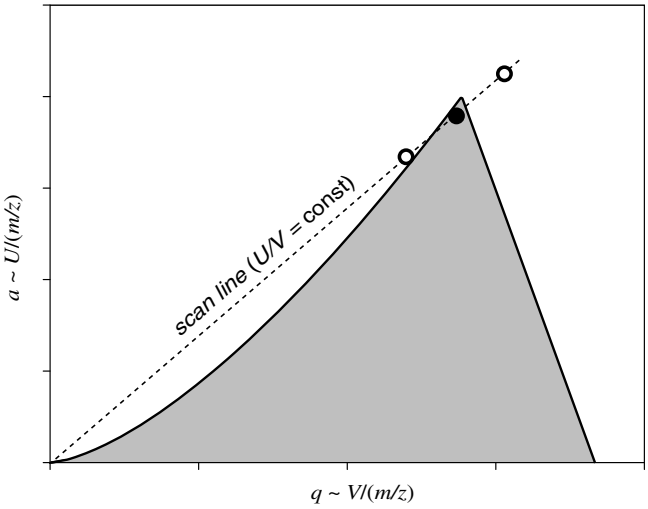


Figure 3.22. A schematic representation of the stability diagram for the quadrupole mass filter depicted in Figure 3.21 (see explanation in the text).

where r_0 is the distance from the central axis of the filter (z-axis) to the surfaces of the electrodes. Combining Eqs. 3-1-1 and 3-5-6 gives rise to complicated equations from which an ion trajectory within the filter can be calculated. Generally, this is a very involved mathematical procedure that is usually carried out by introducing a new variable u, which can represent either x or y:

$$\frac{d^2u}{d^2\xi} = \pm(a+2q\cdot\cos(2\xi))\cdot u \qquad (3\text{-}5\text{-}7)$$

where $\xi = t\omega/2$ and

$$a = a_x = -a_y = \frac{4zeU}{m\omega^2 r_0^2} \quad \text{and} \quad q = q_x = -q_y = \frac{2ze\mathrm{V}}{m\omega^2 r_0^2} \qquad (3\text{-}5\text{-}8)$$

Equation 3-5-7 is known as Mathieu's differential equation. Stable solutions of Mathieu's equation are usually presented as "stability diagrams" in the (*a, q*) coordinates (Fig. 3.22). The highest mass resolving capability will be achieved if the filter is tuned in such a way that the "point" corresponding to the ions of interest is placed close to the apex of the stability region. In this case, a slight increase or decrease of ion *m/z* value will result in an unstable trajectory. To acquire a mass spectrum, both direct current (dc) and rf potentials (U and V) have to be varied, while maintaining their ratio constant. This would bring previously "unstable" ions into the apex region of the stability region, thus allowing their passage through the filter. Another important conclusion from the consideration of the stability diagram is that the quadrupole will become an "ion guide", rather than an ion filter, if the dc component of the electrical field (eq. 3-5-5) is equal to zero, a regime that is commonly referred to as an "rf-only" mode of operation. A very intuitive treatment of the ion motions in quadrupole filters can be found in recent tutorials (135,136).

The *m/z* range of a typical quadrupole MS is limited to 4000. The resolution of a quadrupole mass spectrometer can be adjusted by changing the U/V ratio, the slope of the "scan line" (see Fig. 3.22). Mass resolution is not constant across the *m/z* axis, since the width of the transmission window is "fixed" once the "working" U/V ratio is selected. Typically, mass resolution cannot be increased significantly above the level of several thousand. The scan rate of a typical quadrupole MS is high enough to allow direct coupling to HPLC. MS/MS experiments can be carried out if three quadrupoles are arranged in tandem (a configuration referred to as QqQ, so-called triple quadrupole instruments). The first quadrupole is set to transmit ions of certain *m/z* value (precursor ions), while the second is used as a collision cell. It is operated in the rf-only mode to allow indiscriminate transmission of all ions (precursor and fragments alike) into the third quadrupole, which is scanned to obtain a fragment-ion spectrum. Alternatively, the third quadrupole can be set to allow the transmission of a certain fragment ion *m/z*, while the first one is scanned. Mass spectra acquired in this mode contain peaks of all ions whose fragmentation gives rise to a selected fragment (precursor ion scans). Finally, both first and second quadrupoles can be scanned in concert (maintaining a constant difference), yielding spectra of "constant neutral loss". One of the most-used applications of triple quadrupole mass spectrometers is for quantitation of small molecules. Monitoring a single-fragmentation transition of a single ion *m/z* (single reaction monitoring or SRM) is highly sensitive since none of the electronics are scanning, rather remaining set to transmit a single *m/z*, and is extensively used for quantitation of specific molecules of interest in complex matrices. Multiple reaction monitoring (MRM) is an extension of this method involving rapidly switching (within a few milliseconds) the two transmission quadrupoles between a set of specific *m/z* values for measuring several transitions at once. Quadrupoles are often interfaced with other types of mass analyzers to produce "hybrid" tandem mass spectrometers.

3.5.4. Mass Analyzer As an Ion-Storing Device: The Quadrupole (Paul) Ion Trap and Linear Ion Trap

The idea of using a quadrupolar field to store ions is a logical extension of the concept of a quadrupolar mass filter. It was introduced by Wolfgang Paul, the inventor of the quadrupolar mass filter (137), for which the Nobel Prize in Physics was awarded in 1989. Quadrupole ion traps can be viewed as a linear quadrupole filter that has been "collapsed," so

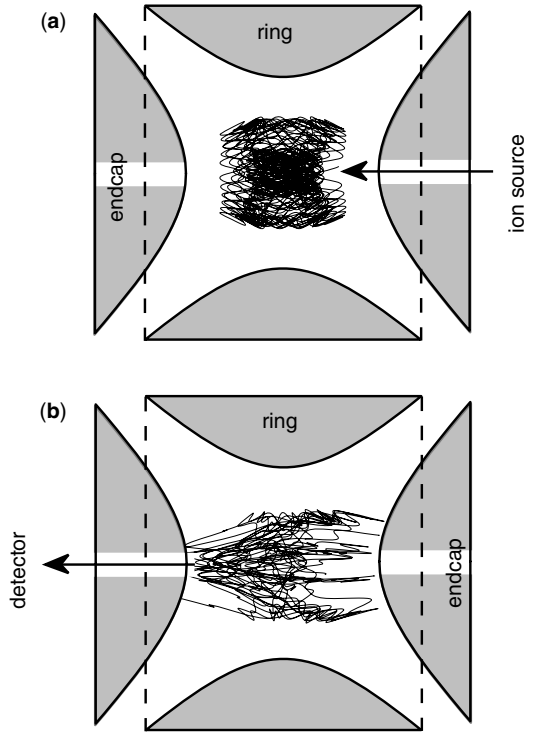


Figure 3.23. A schematic representation of a quadrupolar ion trap. Simulated trajectories of trapped ions (*a*) and ions undergoing resonant excitation (*b*) are shown. [(Courtesy of Prof. Richard W. Vachet, University of Massachusetts-Amherst).]

that the electrical field becomes quadrupolar in all dimensions, not just in the (*x, y*) plane. The confining capacity of a quadrupole ion trap device is due to the formation of a "trapping" potential well when appropriate potentials are applied to three electrodes of hyperbolic cross-sections (two end-caps and one ring electrode, see Fig. 3.23). An oscillating (rf) potential on the ring electrode $\Phi_R = V\cos(\omega t)$ creates a dynamic parabolic (or, more correctly, saddle) field inside the trapping volume, which focuses ions to its center. A potential applied to the end-caps $\Phi_R = U$ is constant (dc), and the field at any point inside the trapping volume is (138,139):

$$\Phi(r,z) = (U - V\sin\omega t)\cdot\left(\frac{r^2-2z^2}{2r_0^2}\right) + \frac{U - V\cos\omega t}{2} \quad (3\text{-}5\text{-}9)$$

where r and z are cylindrical coordinates ($r^2 = x^2 + y^2$), and r_0 is the shortest distance from the center of the trap to the surface of the ring electrode. In practice, the dc potential is applied by providing dc offset to the rf potential, which is applied to the ring electrode. Ion trajectories in such a field will be determined by solutions of Mathieu's equation similar to Eq. 3-5-7. Solutions that correspond to trajectories confined to the trapping volume will form a "stability" region in the (a_z, q_z) plane (Fig. 3.24). Trapping of the externally generated ions that are injected into the trap is

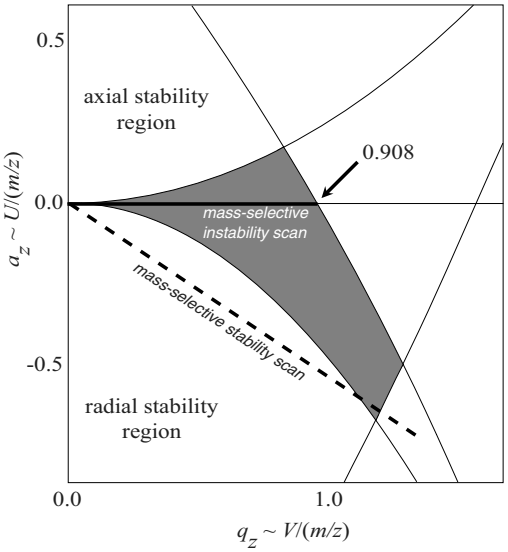


Figure 3.24. A schematic representation of the stability diagram for the quadrupole ion trap depicted in Figure 3.23

facilitated by collisional cooling (using He as a bath gas in the trap at a pressure of $\sim$ 1 mTorr). Consideration of the stability diagram in Figure 3.24 suggests that in order to maximize the "stable" *m/z* range, the ion trap has to be operated at $a_z = 0$. This corresponds to no dc potential applied to the end-caps (the so-called "mass-selective instability mode"). Under such conditions, ions will become ejected from the trap only if their q_z value exceeds 0.908 (Fig. 3.24). This is utilized for the purposes of ion detection in a mass-specific (or, more precisely, *m/z* specific) fashion. Since $q_z = 4zeV/(mr_0^2\omega^2)$, gradual increase of V will result in an increase of q_z and will lead to the ejection of ions of progressively increasing *m/z* values from the trap, followed by their detection.

Trapped (stable) ions of a given *m/z* oscillate at a frequency (known as secular frequency) proportional to ω. If a harmonic potential is applied to the end-caps, resonance conditions will be achieved for those ions whose secular frequency matches that of the applied potential. Resonant absorption of energy by such ions will progressively increase the amplitudes of their oscillations until they become unstable and are ejected from the trapping volume (Fig. 3.23*b*). Resonant excitation can also be used for mass-selective ejection/detection by creating a "hole" in the stability diagram at relatively low q_z values. A gradual increase of V will bring ions of progressively increasing *m/z* values to this hole, making their trajectories unstable and eventually forcing the ions out of the trap.

The ion of interest can also be isolated in the trap using a variety of methods. For example, the ion can be "brought" to the apex of the stability diagram, which would make all other ions unstable (Fig. 3.24). Alternatively, the ion could be "left" on the q_z-axis and the rf amplitude V is then

scanned (increased) to eject all ions with lower *m/z* values. Next, a "resonant hole" is created at higher *m/z* and *V* is scanned to force all high *m/z* ions to exit the trap through this hole. Once the ion of interest has been isolated, resonant excitation can be induced by applying a harmonic potential to the end-caps. The amplitude of the resonant signal can be adjusted such that the ions are not ejected from the trap, but rather undergo a series of collisions with the molecules of the bath gas. If the energy of such collisions is high enough, the ion internal energy will continuously increase and eventually cause ion fragmentation. A scan of rf amplitude *V* after a period of such collisional activation will allow a mass spectrum of fragment ions to be acquired. Precursor ion selection (isolation), collisional activation and fragmentation, as well as mass analysis of the fragment ions all occur sequentially in the same location, hence the term *tandem-in-time* (as opposed *to tandem-in-space*) MS.

Any one of the fragment ions, produced in the course of the MS/MS experiments just described, can be isolated in the trap, activated (the frequency of the resonant potential will have to be adjusted to the new *m/z* value), and fragmented, followed by the acquisition of a mass spectrum of the second generation fragment ions. This process can be repeated any number of times, as long as the number of ions remaining in the trap is high enough to provide a usable S/N. Such experiments are referred to as *multistage tandem MS*, or simply MS^n. Fragmentation efficiency in ion trap MS often approaches 100% for smaller (<1 kDa) peptide ions.

Due to significant improvements in the performance of ion traps, ease of operation, and relatively low cost, these analyzers have recently become very popular, both as standalone mass spectrometers and as a part of hybrid instruments. Fast scan time allows ion traps to be coupled directly to HPLC (using ESI as an interface). Ion traps are also well suited for pulsed ionization sources (e.g., MALDI), although the *m/z* range of most commercial ion trap instruments is limited to 6000 (with optimal performance <2500), which limits the scope of biomolecules amenable to analysis. Utilization of ESI can extend the range of biopolymers amenable to analysis (because of multiple charging, ESI generated protein ions usually fall within the "allowed" *m/z* range). However, one needs to be aware of another potential problem, which is related to the space-charge effect. Since the trapping volume is usually very small, accumulation of a large number of charges, which may exceed the number of trapped proteins by more than an order of magnitude due to multiple charging, may give rise to strong electrical fields inside the trap. Space-charging phenomena have a detrimental effect on sensitivity and resolution, and often result in mass spectral artifacts, such as shifts in measured *m/z* or ghost peaks (140,141). More information on the general principles of quadrupole ion traps can be found in several recent review and tutorial articles (138,139,142,143).

A significant limitation of the three-dimensional (3D) ion trap, as mentioned above, is the small trapping volume, resulting in a limit to the number of ions (charges) that can be accumulated in a single trapping cycle without adverse effects (e.g., space-charge distortion). For this reason, the linear ion trap, with its much larger trapping capacity, recently has become a popular device that provides the same versatility, while avoiding the limitations imposed by space charging. The concept of the linear trap was first described in the 1960s with research on the quadrupole storage ring by Drees and Paul (144). The simplest design of a linear trap is a segmented quadrupole (based on a triple quadrupole design) in which the central pressurized segment confines the ions radially in a rf quadrupolar field, while the terminal segments provide repulsive dc potentials at either end that contain the ions within the trap. There are many advantages to this design over the 3D trap, the main one being a much higher trapping efficiency (theoretically up to 100% in a linear trap versus 5–10% in a 3D ion trap) due to the lack of an axial quadrupolar field that would otherwise tend to repel ions. Additionally, the larger internal volume of the linear trap can accommodate more charges that can be distributed along the length of the quadrupolar trapping field, thus minimizing space-charge effects (145,146).

Tandem (MS^n) experiments can be performed in a linear trap in a manner similar to a 3D ion trap, by using resonance excitation of the ion of interest in the trapping quadrupole (147). Ion-detection is normally performed by resonance ejection of ions radially through slots in the quadrupole rods to be detected using standard ion detection methods. Alternatively, the ions can be ejected axially along the length of the quadrupole and transmitted to further components of the mass spectrometer. This has led to a variety of uses of linear traps in hybrid mass spectrometers. For instance, fragment ions could be transmitted to a Fourier transform–ion cyclotron resonance (FT ICR) or Orbitrap (see below) cell for high-resolution detection, as will be discussed in Section 3.5.8. Generally this results in a significant increase in duty cycle since ions can be trapped, fragmented, and stored in the linear trap region while other measurements are being performed elsewhere in these hybrid instruments.

3.5.5. Mass Analyzer As an Ion Storing Device: FT ICR MS

Ion confinement to a limited volume can also be achieved by using a combination of electrostatic and magnetic fields. The simplest configuration of such a trapping device is depicted in Figure 3.25. A dc potential applied to the front and back plates will restrict the ion motion in the *z*-axial direction, while a constant magnetic field in the same direction will induce a circular (cyclotron) motion in the (*x, y*) plane. In accordance with Eq. 3-1-1, the centripetal force will be equal to the Lorentz force exerted on the ion by

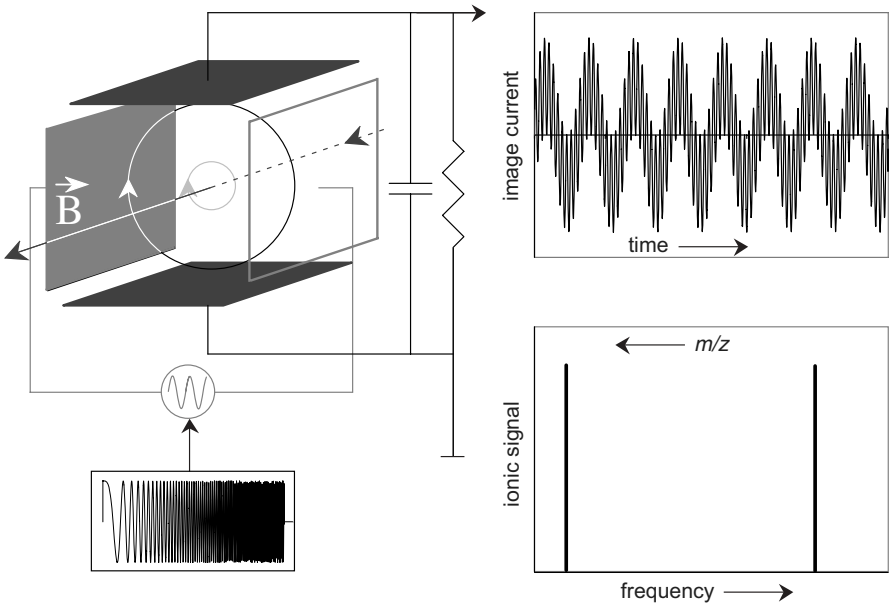


Figure 3.25. Principle of ion trapping, broad-band excitation and detection in FT ICR MS

the magnetic field, and so the cyclotron motion frequency ω_c will be uniquely determined by the magnetic field strength B and the *m/z* ratio of the ion:

$$\omega_c = \frac{zeB}{m} \qquad (3\text{-}5\text{-}10)$$

Therefore, measuring ω_c can be used for mass analysis. Various modifications of this concept were brought forward in the late 1940s – early 1950s under the names *magnetic* TOF mass spectrometer, where the flight time was actually the cyclotron motion period $T = 2\pi/\omega$ (148,149), *magnetic period* mass spectrometer or *mass synchrometer* (150,151), *omegatron* (152,153) and *magnetic resonance* mass spectrometer (154). Since frequency is a physical parameter that can be measured very accurately, mass spectrometers based on the principle of cyclotron motion can provide the highest accuracy *m/z* measurements. However, it was not until the mid-1970s that ICR MS (155) finally became the powerful analytical technique we know now due to the introduction of digital FT as a means of producing the mass spectrum (156). A more detailed historical account of the development of ICR MS can be found in an excellent review by Marshall (157).

Ion detection in FT ICR MS is done by measuring the magnitude of the "image current" induced on the plates by the ion orbiting between them (Fig. 3.25). Unsynchronized motion of a large number of ions will result in no net current due to the random distribution of phases among the ion population. Therefore, ion detection must be preceded by ion excitation (e.g., by applying a uniform harmonic electric field in the direction orthogonal to the magnetic field). If the field frequency is the same as the cyclotron frequency of the orbiting ions, they will be synchronized (brought in phase with the field). Furthermore, such resonant conditions will elevate ion kinetic energy, increasing the radii of their orbits and, therefore, the magnitude of the image current induced by each ion. A homogeneous population of ions whose orbiting motions are in phase will induce an image current on the "receiver" plates (the top and bottom plate on Fig. 3.25) of the form $I = I_0 \cdot \sin(\omega t + \alpha)$, whose angular frequency ω will be equal to the cyclotron frequency ω_c of the orbiting ions. The amplitude of the current I_0 will be proportional to the number of ions in the population and essentially independent of the cyclotron frequency ω_c under typical ICR conditions (158). If several ion populations (of different *m/z* ratios) are present in the cell and are excited to higher orbits, the resulting image current will be a superposition of several sinusoidal signals, each having a characteristic angular frequency uniquely determined by the magnetic field strength and the *m/z* ratio of the corresponding ion population (Eq. 3-5-10). [The actual cyclotron frequency in a real ICR cell will be lower than ω_c due to the presence of an electrostatic potential, which traps the ions in the axial (z) direction.] Fourier transformation of such a spectrum (from the time domain to the frequency domain) will allow the cyclotron frequency of each ion population to be determined and *m/z* values calculated (Fig. 3.26).

Fourier transform ICR MS offers the advantage, over other MS detection methods, of being able to detect all ions simultaneously across a wide *m/z* range within a very short period of time. Increasing the measurement time in an FT ICR MS experiment enhances not only the S/N, but also the resolution. The latter is usually limited by "collisional

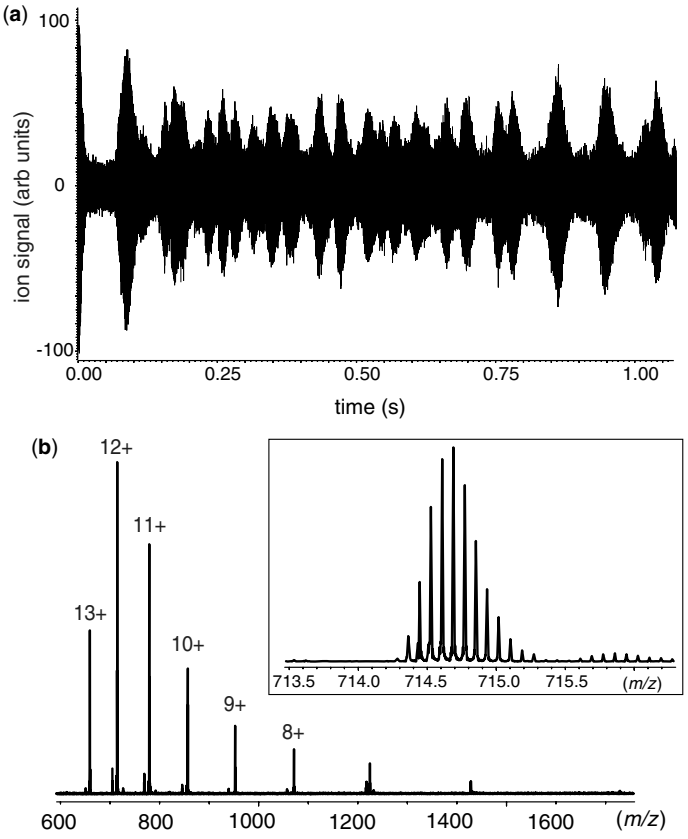


Figure 3.26. The FT ICR mass spectra of ubiquitin: raw data in the time domain (*a*) and the FT spectrum (frequency domain) (*b*).

damping" of the cyclotron motion (in contrast to quadrupole ion traps, ion-neutral collisions in the ICR cell have detrimental effects on the quality of mass spectra). Such collisions result in broadening and shape distortion of the ion peaks (without altering the *m/z* values), hence the very high vacuum requirements for high-resolution measurements (more demanding compared to most other mass analyzers). Due to the limited volume of the ICR cell, space-charge effects also negatively impact the quality of FT ICR MS data (causing systematic shifts in the observed cyclotron frequencies of ions and, therefore, limiting the accuracy of mass measurements). Since all ions experience the same space-charge induced frequency shift, accurate mass measurements can be carried out using internal calibration (159). Commercial FT ICR instruments offer resolution exceeding 100,000 in the broadband mode, far outperforming most other types of mass analyzer (except Orbitrap, see Section 3.5.6). Image current detection is generally less sensitive compared to the "classical" ion counting techniques utilized by other types of mass analyzers. However, FT ICR MS detection is nondestructive and the data acquisition can be carried out with the same ion population over an extended period of time using multiple remeasurements (160,161). In fact, it is possible to use FT ICR MS to trap and detect individual multiply charged macromolecular ions (162).

Inverse FT (from the frequency to the time domain) allows "custom" excitation spectra to be designed. Such *stored waveform inverse Fourier transform* (SWIFT) excitation is used for highly selective ion excitation and/or ejection from the trap. This excitation is an extremely effective method for ion isolation in the ICR cell. Fragmentation of the isolated ions can be induced by a variety of means, including collisional activation (e.g., *sustained off-resonance irradiation*, SORI CID), photoactivation (*e.g.*, *infrared multi-photon dissociation*, IRMPD (105)), electron capture, ECD (111), and so on. Most of these methods (with the exception of ECD) are slow-heating methods according to McLuckey's classification (91). Higher-energy CAD can be implemented by using collisional activation of ions in the external source (e.g., nozzle-skimmer CAD in the ESI interface). Trapping the activated ions [either in the ICR cell or in the external reservoir, e.g., a hexapole ion guide (163)] allows the yield of the dissociation to be increased dramatically compared to other mass analyzers. Since FT ICR MS is a trapping device, it allows multiple stages of ion fragmentation to be carried out in sequence. Combination of several ion fragmentation techniques in one experiment often provides significant improvement of the sequence coverage of macromolecular ions (164). More information on general principles of FT ICR MS and its applications to biomolecular analyses can be found in recent reviews (165,166).

3.5.6. Mass Analyzer as An Ion Storing Device: Orbitrap MS

A relative newcomer to the battery of mass analyzers was recently developed by Makarov (167,168), namely, the Orbitrap. The concept is similar to the Kingdon trap, developed in 1923, that relies only on electrostatic fields to trap the ions in a cylindrical outer electrode around a thin wire (169). The Orbitrap design is illustrated in Fig. 3.27. Basically, ions are injected into a "quadrologarithmic" electrostatic potential created by a central spindle electrode surrounded by the curved outer electrode, described by

$$U(r,z) = \frac{k}{2}\left(z^2 - \frac{r^2}{2}\right) + \frac{k}{2}\cdot(R_m)^2\cdot\ln\left[\frac{r}{R_m}\right] + C \quad (3\text{-}5\text{-}11)$$

where r and z are cylindrical coordinates within the electrode, C is a constant, R_m is the characteristic radius, and k is the curvature of the field. Once ions are injected into this electrostatic field with a velocity perpendicular to the z-axis of the trap and with an offset from the center of the trap, they orbit around the central electrode, but also oscillate along the z-axis with a frequency simply related to their m/z

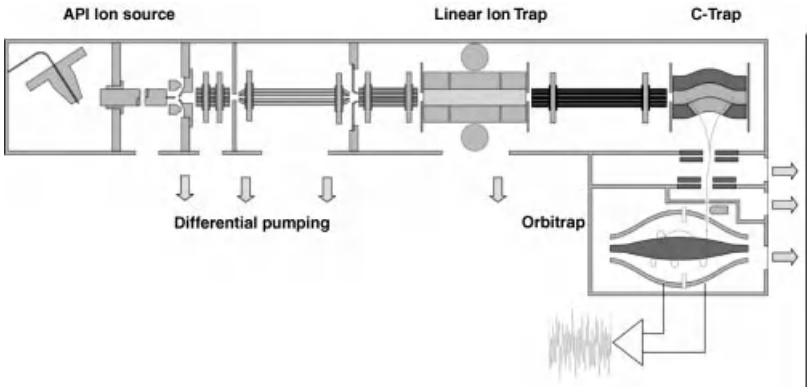


Figure 3.27. Schematic diagram of an Orbitrap mass analyzer interfaced to a linear ion trap. [Reprinted with permission from (170). Copyright © 2005 John Wiley & Sons.]

and the field curvature (167):

$$\omega = \sqrt{(z/m) \cdot k} \qquad (3\text{-}5\text{-}12)$$

Ions can then be detected based on their frequencies of oscillation by image current in a method analogous to the FT ICR analyzer described above, with the time domain signal converted to a mass spectrum by FT. Because the frequency along the *z*-axis is essentially independent of the energy and spatial arrangement of the ions, very high resolution and high mass accuracy measurements can be achieved by this method of detection (171). The Orbitrap analyzer has many advantages, including high resolution and high mass accuracy due to the static nature of the well-defined electric field in the trap (cf. FT ICR). However, the higher capacity and independence of *m/z* on the trapping ability relative to other trapping devices means that the Orbitrap can achieve a much higher dynamic range. Also, since the frequency resolving power is inversely proportional to $(m/z)^{1/2}$ (Eq. 3-5-12) rather than *m/z* (Eq. 3-5-10), the usable high-resolution mass range is significantly enhanced for Orbitrap versus FT ICR. Resolution is limited by the residence time of ions in the trap, however; nevertheless, an Orbitrap instrument can routinely achieve 100,000 resolving power and achieves a higher dynamic and *m/z* range with generally much less tuning required. Coupled to the fact that an Orbitrap is significantly smaller and does not require cryogenic cooling, this has become an attractive device in modern laboratories limited by space and consumables budgets, but who still require high-resolution MS capabilities.

3.5.7. Ion-Mobility Analyzers

Another technique that has enjoyed significant attention in recent years relies on the idea that macromolecular ions with different structure will pass through a drift tube filled with an intermediate pressure of neutral gas molecules at different rates depending on their collisional cross-sections. First developed in the 1950s by McDaniel et al. (172) and Kebarle et al. (173) for the study of ion-molecule reactions in gases (172,173), drift time ion-mobility spectrometry measures the amount of time that an ion takes to migrate through a buffer gas in the presence of a low electric field. Under conditions where diffusion processes are dominant (thermal energy from collisions exceeds energy of the ions due to the electric field), the velocity, *v* of the ion under the influence of electric field *E,* is inversely proportional to its collisional cross-section (Ω):

$$v = \frac{3ze}{16N\Omega}\frac{1}{\left(\dfrac{2\pi}{\mu k_B T}\right)^{1/2}} \cdot E \qquad (3\text{-}5\text{-}13)$$

where *z* is the charge on the ion, μ is the reduced mass of the ion and the buffer gas, k_B is the Boltzmann constant, and *N* is the number density of the buffer gas. Thus, measuring the drift time of an ion through a drift cell of known length yields characteristic information that has been used for the analysis of small molecules (e.g., drugs and explosives). Coupled to a mass spectrometer, it now becomes possible to also measure *m/z* of the ions, thus enabling calculation of the collisional cross-section.

Seminal work by Bowers, Clemmer, Jarrold, and their co-workers with peptides and small proteins using in-house built instrumentation laid the groundwork for understanding how ion mobility can be used to determine protein conformations in the gas phase. (174–178). Modeling based on molecular dynamics simulations can be used to fit the measured ion mobilities to reasonable molecular structures that fit with the observed collisional cross-sections. However, only recently has commercial instrumentation become available to enable more widespread usage. Since then many studies have been performed that suggest proteins can retain structures in the gas phase that reflect their solution properties (179,180), and even macromolecular complexes can, to a large extent, be maintained (181). This technique also appears to complement, in many ways, the analysis of

charge-state distributions of macromolecular ions electrosprayed from solution, as described in Section 4.1.4. A complete description of ion mobility mass spectrometry (IM–MS) for the study of biomolecular structure is beyond the scope of this text, but there are some excellent recent reviews (182,183).

3.5.8. Hybrid Mass Spectrometers

In addition to the mass analyzers described in the preceding sections, there is an increasing number of *hybrid* mass spectrometers that combine more than one type of mass analyzer in a single instrument. Hybrid mass spectrometers often feature enhanced performance as a result of capitalizing on the advantages provided by each component. While earlier versions of hybrid MS systems almost always utilized a sector instrument as the first stage, more recently the focus has been shifting toward TOF MS and quadrupole ion trap MS as the core component. Combination of a quadrupole mass filter and a reflectron-TOF mass analyzer has been particularly popular (184). This configuration offers tandem capabilities, high resolution, an extended *m/z* range, and fast analysis time. Another hybrid instrument that has gained popularity in recent years combines a quadrupole ion trap and a TOF mass analyzer (185). Also, two recent high-end commercial products combine a linear ion trap (186) or a quadrupole filter (187) with FT ICR MS for high-resolution detection.

Finally, an instrument that has enjoyed immense popularity in recent years is the hybrid linear trap–Orbitrap mass spectrometer. This combines the speed of acquisition and MS^n capabilities of a linear trap with the high-resolution and high-mass accuracy of an Orbitrap analyzer. This system can acquire multiple MS and MS/MS spectra per second, compatible with LC separations, but is also able to pass user-selected (or software-selected based on criteria) ions to the Orbitrap for accurate mass determination. This combination has proven exceptionally useful in the field of proteomics where rapid MS/MS sequence data are complemented by high-resolution parent ion mass determination, giving increased confidence in peptide mass assignments (170,188,189).

This chapter, gave a brief overview of the concepts underlying MS and the technology available to achieve mass measurements. Chapters 4–9 describe some of the many and increasing number of applications of these MS methods in the biophysical arena.

REFERENCES

1. Thomson, J. J. (1913) *Rays of positive electricity and their application to chemical analyses*, Longmans Green and Co., London, New York.
2. Griffiths, I. W. (1997) J. J. Thomson—The centenary of his discovery of the electron and of his invention of mass spectrometry, *Rapid Commun. Mass Spectrom. 11*, 3–16.
3. Aston, F. W. (1920) Isotopes and atomic weights, *Nature (London) 105*, 617–619.
4. Audi, G. and Wapstra, A. H. (1993) The 1993 Atomic Mass Evaluation 1. Atomic Mass Table, *Nuclear Phys. A 565*, 1–65.
5. Coplen, T. B., Bohlke, J. K., De Bievre, P., Ding, T., Holden, N. E., Hopple, J. A., Krouse, H. R., Lamberty, A., Peiser, H. S., Revesz, K., Rieder, S. E., Rosman, K. J. R., Roth, E., Taylor, P. D. P., Vocke, R. D. and Xiao, Y. K. (2002) Isotope-abundance variations of selected elements - (IUPAC Technical Report), *Pure Appl. Chem. 74*, 1987–2017.
6. Vocke, R. D. (1999) Atomic weights of the elements 1997 (technical report), *Pure Appl. Chem. 71*, 1593–1607.
7. Rosman, K. J. R. and Taylor, P. D. P. (1999) Isotopic compositions of the elements 1997, *pure Appl. chem 70*, 217–235.
8. Yergey, J. A. (1983) A general approach to calculating isotopic distributions for mass spectrometry, *Int. J. Mass Spectrom. Ion Proc. 52*, 337–349.
9. Hsu, C. S. (1984) Diophantine approach to isotopic abundance calculations, *Anal. Chem. 56*, 1356–1361.
10. Yergey, J., Heller, D., Hansen, G., Cotter, R. J. and Fenselau, C. (1983) Isotopic distributions in mass spectra of large molecules, *Anal. Chem. 55*, 353–356.
11. Kubinyi, H. (1991) Calculation of isotope distributions in mass spectrometry - a trivial solution for a nontrivial problem, *Analyt. Chim. Acta 247*, 107–119.
12. Rockwood, A. L., Vanorden, S. L. and Smith, R. D. (1995) Rapid calculation of isotope distributions, *Anal. Chem. 67*, 2699–2704.
13. Rockwood, A. L. and VanOrden, S. L. (1996) Ultrahigh-speed calculation of isotope distributions, *Anal. Chem. 68*, 2027–2030.
14. Schutz, C. N. and Warshel, A. (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models?, *Proteins 44*, 400–417.
15. Prâokai, L. (1990) *Field desorption mass spectrometry*, M. Dekker, New York.
16. Macfarlane, R. D. and Torgerson, D. F. (1976) Californium-252 plasma desorption mass spectroscopy, *Science 191*, 920–925.
17. Barber, M., Bordoli, R. S., Sedgwick, R. D. and Tyler, A. N. (1981) Fast atom bombardment of solids (FAB)—a new ion source for mass spectrometry., *J. Chem. Soc. Chem. Comm. 1981*, 325–327.
18. Henry, C. (1997) FAB MS: still FABulous?, *Anal. Chem. 69*, 625A–627A.
19. Seifert, W. E., Jr. and Caprioli, R. M. (1996) Fast atom bombardment mass spectrometry, *Methods Enzymol. 270*, 453–486.
20. Das, P. R. and Pramanik, B. N. (1998) Fast atom bombardment mass spectrometric characterization of peptides, *Mol. Biotechnol. 9*, 141–154.

21. Chapman, J. R. (1993) *Practical organic mass spectrometry: a guide for chemical and biochemical analysis*, 2nd ed., John Wiley & sons, Inc., Chichester; New York.
22. Watson, J. T. (1997) *Introduction to mass spectrometry*, 3rd ed., Lippincott–Raven, Philadelphia.
23. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules., *Science 246*, 64–71.
24. Fenn, J. B. (2003) Electrospray wings for molecular elephants (Nobel lecture), *Angew. Chem. Int. Ed. Engl. 42*, 3871–3894.
25. Chapman, S. (1937) Carrier mobility spectra of spray electrified liquids., *Phys. Rev. 52*, 184–190.
26. Rayleigh, J. W. S. (1882) On the equilibrium of liquid conducting masses charged with electricity., *Philos. Mag. 14*, 184–186.
27. Gilbert, W. (1600) *De magnete, magneticisque corporibus, et magno tellure magnete physiologia nova, plurimis & argumentis, & experimentis demonstrata.*, P. Short, London.
28. Dole, M., Cox, H. L. and Gieniec, J. (1973) Electrospray mass spectroscopy., *Adv. Chem. Ser. 125*, 73–84.
29. Dole, M., Hines, R. L., Mack, L. L., Mobley, R. C., Ferguson, L. D. and Alice, M. B. (1968) Gas phase macroions., *Macromolecules 1*, 96–97.
30. Richardson, M. J. (1964) The direct observation of polymer molecules and determination of their molecular weight., *Proc. R. Soc. Lond A 279*, 50–61.
31. Dole, M., Mack, L. L. and Hines, R. L. (1968) Molecular beams of macroions., *J. Chem. Phys. 49*, 2240–2249.
32. Mack, L. L., Kralik, P., Rheude, A. and Dole, M. (1970) Molecular beams of macroions. 2., *J. Chem. Phys. 52*, 4977–4986.
33. Thomson, B. A., Iribarne, J. V. and Dzledzic, P. J. (1982) Liquid ion evaporation/mass spectrometry/mass spectrometry for the detection of polar and labile molecules., *Anal. Chem. 54*, 2219–2224.
34. Iribarne, J. V., Dzledzic, P. J. and Thomson, B. A. (1983) Atmospheric pressure ion evaporation mass spectrometry., *Int. J. Mass Spectrom. Ion Proc. 50*, 331–347.
35. Alexandrov, M. L., Gall, L. N., Krasnov, N. V., Nikolaev, V. I., Pavlenko, V. A. and Shkurov, V. A. (1984) Ion extraction from solutions at atmospheric pressure - a method of mass spectrometric analysis of bioorganic substances., *Dokl. Acad. Nauk SSSR 277*, 379–383.
36. Yamashita, M. and Fenn, J. B. (1984) Negative ion production with the electrospray ion source., *J. Phys. Chem. 88*, 4671–4675.
37. Alexandrov, M. L., Baram, G. I., Gall, L. N., Krasnov, N. V., Kusner, Y. S., Mirgorodskaya, O. A., Nikolaev, V. I. and Shkurov, V. A. (1985) Formation of beams of quasimolecular ions of peptides from solutions., *Bioorg. Khim. 11*, 700–704.
38. Alexandrov, M. L., Baram, G. I., Gall, L. N., Grachev, M. A., Knorre, V. D., Krasnov, N. V., Kusner, Y. S., Mirgorodskaya, O. A., Nikolaev, V. I. and Shkurov, V. A. (1985) Application of a novel mass spectrometric method to sequencing of peptides., *Bioorg. Khim. 11*, 705–708.
39. Alexandrov, M. L., Gall, L. N., Krasnov, N. V., Nikolaev, V. I., Pavlenko, V. A., Shkurov, V. A., Baram, G. I., Grachev, M. A., Knorre, V. D. and Kusner, Y. S. (1984) Direct coupling of a microcolumn liquid chromatograph and a mass-spectrometer, *Bioorg. Khim. (USSR) 10*, 710–712.
40. Whitehouse, C. W., Dreyer, R. N., Yamashita, M. and Fenn, J. B. (1985) Electrospray interface for liquid chromatographs and mass spectrometers., *Anal. Chem. 57*, 675–679.
41. Vestal, M. L. (2001) Methods of ion generation, *Chem. Rev. 101*, 361–375.
42. Meng, C. K., Mann, M. and Fenn, J. B. (1988) Of protons or proteins., *Z. Phys. D 10*, 361–368.
43. Wong, S. F., Meng, C. K. and Fenn, J. B. (1988) Multiple charging in electrospray ionization of polyethylene glycols, *J. Phys. Chem. 92*, 546–550.
44. Mann, M., Meng, C. K. and Fenn, J. B. (1989) Interpreting mass-spectra of multiply charged ions, *Anal. Chem. 61*, 1702–1708.
45. Ferrige, A. G., Seddon, M. J. and Jarvis, S. (1991) Maximum entropy deconvolution in electrospray mass spectrometry, *Rapid Commun. Mass Spectrom. 5*, 374–377.
46. Reinhold, B. B. and Reinhold, V. N. (1992) Electrospray ionization mass spectrometry - deconvolution by an entropy based algorithm, *J. Am. Soc. Mass Spectrom. 3*, 207–215.
47. Zhang, Z. Q., Guan, S. H. and Marshall, A. G. (1997) Enhancement of the effective resolution of mass spectra of high-mass biomolecules by maximum entropy-based deconvolution to eliminate the isotopic natural abundance distribution, *J. Am. Soc. Mass Spectrom. 8*, 659–670.
48. Zhang, Z. Q. and Marshall, A. G. (1998) A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra, *J. Am. Soc. Mass Spectrom. 9*, 225–233.
49. Horn, D. M., Zubarev, R. A. and McLafferty, F. W. (2000) Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules, *J. Am. Soc. Mass Spectrom. 11*, 320–332.
50. Pramanik, B. N., Ganguly, A. K. and Gross, M. L. (2002) *Applied electrospray mass spectrometry*, Marcel Dekker, New York.
51. Cole, R. B. (2010) *Electrospray and MALDI mass spectrometry: fundamentals, instrumentation, practicalities and biological applications*, 2nd ed., Wiley, Hoboken, N.J.
52. Wilm, M. S. and Mann, M. (1994) Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last?, *Int. J. Mass Spectrom. Ion Proc. 136*, 167–180.
53. Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source, *Anal. Chem. 68*, 1–8.
54. Juraschek, R., Dulcks, T. and Karas, M. (1999) Nanoelectrospray - more than just a minimized-flow electrospray ionization source, *J. Am. Soc. Mass Spectrom. 10*, 300–308.
55. Karas, M., Bahr, U. and Dulcks, T. (2000) Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine, *Fresen. J. Anal. Chem. 366*, 669–676.

56. Conzemius, R. J. and Capellen, J. M. (1980) A review of the applications to solids of the laser ion-source in mass-spectrometry, *Int. J. Mass Spectrom. Ion Proc 34*, 197–271.
57. Tanaka, K., Ido, Y., Akita, S., Yoshida, Y. and Yoshida, T. (1987) Detection of high mass molecules by laser desorption time-of-flight mass spectrometry, In *2-d Japan-China Joint Symposium on Mass Spectrometry*, pp 185–188, Bando Press, Osaka, Japan.
58. Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem. 60*, 2299–2301.
59. Tanaka, K. (2003) The origin of macromolecule ionization by laser irradiation (Nobel lecture), *Angew. Chem. Int. Ed. Engl. 42*, 3860–3870.
60. Chiang, C. K., Chen, W. T. and Chang, H. T. (2011) Nanoparticle-based mass spectrometry for the analysis of biomolecules, *Chem. Soc. Rev. 40*, 1269–1281.
61. Nelson, R. W., Rainbow, M. J., Lohr, D. E. and Williams, P. (1989) Volatilization of high molecular weight DNA by pulsed laser ablation of frozen aqueous solutions, *Science 246*, 1585–1587.
62. Overberg, A., Karas, M., Bahr, U., Kaufmann, R. and Hillenkamp, F. (1990) Matrix-assisted infrared laser (2.94 μm) desorption ionization mass-spectrometry of large biomolecules, *Rapid Commun. Mass Spectrom. 4*, 293–296.
63. Dreisewerd, K., Berkenkamp, S., Leisner, A., Rohlfing, A. and Menzel, C. (2003) Fundamentals of matrix-assisted laser desorption/ionization mass spectrometry with pulsed infrared lasers, *Int. J. Mass Spectrom. 226*, 189–209.
64. Charvat, A. and Abel, B. (2007) How to make big molecules fly out of liquid water: applications, features and physics of laser assisted liquid phase dispersion mass spectrometry, *Phys. Chem. Chem. Phys. 9*, 3335–3360.
65. Moyer, S. C. and Cotter, R. J. (2002) Atmospheric pressure MALDI, *Anal. Chem. 74*, 468A–476A.
66. Von Seggern, C. E., Moyer, S. C. and Cotter, R. J. (2003) Liquid infrared atmospheric pressure matrix-assisted laser desorption/ionization ion trap mass spectrometry of sialylated carbohydrates, *Anal. Chem. 75*, 3212–3218.
67. Laiko, V. V., Taranenko, N. I., Berkout, V. D., Yakshin, M. A., Prasad, C. R., Lee, H. S. and Doroshenko, V. M. (2002) Desorption/ionization of biomolecules from aqueous solutions at atmospheric pressure using an infrared laser at 3 μm, *J. Am. Soc. Mass Spectrom. 13*, 354–361.
68. Laiko, V. V., Taranenko, N. I. and Doroshenko, V. M. (2006) On the mechanism of ion formation from the aqueous solutions irradiated with 3 μm IR laser pulses under atmospheric pressure, *J. Mass Spectrom. 41*, 1315–1321.
69. Marshall, A. G., Senko, M. W., Li, W., Li, M., Dillon, S., Guan, S. and Logan, T. M. (1997) Protein molecular mass to 1 Da by ^{13}C, ^{15}N double-depletion and FT ICR mass spectrometry, *J. Am. Chem. Soc. 119*, 433–434.
70. Kolli, V. S. and Orlando, R. (1997) A new strategy for MALDI on magnetic sector mass spectrometers with point detectors, *Anal. Chem. 69*, 327–332.
71. Tomer, K. B. (2001) Separations combined with mass spectrometry, *Chem. Rev. 101*, 297–328.
72. Gelpi, E. (2002) Interfaces for coupled liquid phase separation/mass spectrometry techniques. An update on recent developments, *J. Mass Spectrom. 37*, 241–253.
73. Romijn, E. P., Krijgsveld, J. and Heck, A. J. R. (2003) Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics, *J. Chromatogr. A 1000*, 589–608.
74. Tomlinson, A. J. and Chicz, R. M. (2003) Microcapillary liquid chromatography/tandem mass spectrometry using alkaline pH mobile phases and positive ion detection, *Rapid Commun. Mass Spectrom. 17*, 909–916.
75. Kennedy, R. T., German, I., Thompson, J. E. and Witowski, S. R. (1999) Fast analytical-scale separations by capillary electrophoresis and liquid chromatography, *Chem. Rev. 99*, 3081–3132.
76. Churchwell, M. I., Twaddle, N. C., Meeker, L. R. and Doerge, D. R. (2005) Improving LC–MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS, *J. Chromatogr. B-Analyti. Technol. Biomed Life Sci. 825*, 134–143.
77. McMaster, M. (2005) *LC/MS: A Practical User's Guide*, Wiley Interscience, Hoboken, NJ.
78. Cappiello, A.,(Ed.) (2006) *Advances in LC–MS Instrumentation*, Elsevier, Amsterdam.
79. Maxwell, E. J. and Chen, D. D. Y. (2008) Twenty years of interface development for capillary electrophoresis–electrospray ionization-mass spectrometry, *Anal. Chim. Acta 627*, 25–33.
80. Hommerson, P., Khan, A. M., de Jong, G. J. and Somsen, G. W. (2011) Ionization techniques in capillary electrophoresis-mass spectrometry: Principles, design and application, *Mass Spectrom. Rev. 30*, 1096–1120
81. Nesbitt, C. A., Zhang, H. X. and Yeung, K. K. C. (2008) Recent applications of capillary electrophoresis-mass spectrometry (CE–MS): CE performing functions beyond separation, *Anal. Chim. Acta 627*, 3–24.
82. Gaberc-Porekar, V. and Menart, V. (2001) Perspectives of immobilized-metal affinity chromatography, *J. Biochem. Biophys. Methods 49*, 335–360.
83. Lee, W.-C. and Lee, K. H. (2004) Applications of affinity chromatography in proteomics, *Anal. Biochem. 324*, 1–10.
84. Dixon, S. P., Pitfield, I. D. and Perrett, D. (2006) Comprehensive multi-dimensional liquid chromatographic separation in biomedical and pharmaceutical analysis: a review, *Biomed. Chromatogr. 20*, 508–529.
85. Tang, J., Gao, M. X., Deng, C. H. and Zhang, X. M. (2008) Recent development of multi-dimensional chromatography strategies in proteome research, *J. Chromatogr. B-Analyt. Technol. Biomed. and Life Sci. 866*, 123–132.
86. Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure, *Biomed. Environ. Mass Spectrom. 16*, 99–111.

87. Biemann, K. (1990) Appendix 5. Nomenclature for peptide fragment ions (positive ions), *Methods Enzymol. 193*, 886–887.
88. Ngoka, L. C. and Gross, M. L. (1999) A nomenclature system for labeling cyclic peptide fragments, *J. Am. Soc. Mass Spectrom. 10*, 360–363.
89. McLuckey, S. A. and Wells, J. M. (2001) Mass analysis at the advent of the 21st century, *Chem. Rev. 101*, 571–606.
90. Jennings, K. R. (2000) The changing impact of the collision-induced decomposition of ions on mass spectrometry, *Int. J. Mass Spectrom. 200*, 479–493.
91. McLuckey, S. A. and Goeringer, D. E. (1997) Slow heating methods in tandem mass spectrometry, *J. Mass Spectrom. 32*, 461–474.
92. Dunbar, R. C. (1992) Infrared radiative cooling of gas-phase ions, *Mass Spectrom. Rev. 11*, 309–339.
93. Fabris, D., Michele Kelly, Constance Murphy, Zhuchun Wu and Catherine Fenselau. (1993) High-energy collision-induced dissociation of multiply charged polypeptides produced by electrospray, *J. Am. Soc. Mass Spectrom. 4*, 652–661.
94. Kolli, V. S. K. and Orlando, R. (1995) Complete Sequence Confirmation of Large Peptides by High Energy Collisional Activation of Multiply Protonated Ions, *J. Am. Soc. Mass Spectrom. 6*, 234–241.
95. Johnson, R. S., Martin, S. A. and Biemann, K. (1988) Collision-induced fragmentation of $(M+H)^+$ ions of peptides. Side chain specific sequence ions, *Int. J. Mass Spectrom. Ion Proc. 86*, 137–154.
96. Bowie, J. H., Brinkworth, C. S. and Dua, S. (2002) Collision-induced fragmentations of the $(M-H)^-$ parent anions of underivatized peptides: an aid to structure determination and some unusual negative ion cleavages, *Mass Spectrom. Rev. 21*, 87–107.
97. Bilusich, D., Brinkworth, C. S. and Bowie, J. H. (2004) Negative ion mass spectra of Cys-containing peptides. The characteristic Cys gamma backbone cleavage: a joint experimental and theoretical study, *Rapid Commun. Mass Spectrom. 18*, 544–552.
98. Bilusich, D. and Bowie, J. H. (2007) Identification of intermolecular disulfide linkages in underivatised peptides using negative ion electrospray mass spectrometry. A joint experimental and theoretical study, *Rapid Commun. Mass Spectrom. 21*, 619–628.
99. Chrisman, P. A. and McLuckey, S. A. (2002) Dissociations of disulfide-linked gaseous polypeptide/protein anions: ion chemistry with implications for protein identification and characterization, *J. Proteome Res. 1*, 549–557.
100. Zhang, M. and Kaltashov, I. A. (2006) Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection, *Anal. Chem. 78*, 4820–4829.
101. Andreazza, H. J. and Bowie, J. H. (2010) The application of negative ion electrospray mass spectrometry for the sequencing of underivatized disulfide-containing proteins: insulin and lysozyme, *Physical chemistry chemical physics: PCCP 12*, 13400–13407.
102. Dongre, A. R., Somogyi, A. and Wysocki, V. H. (1996) Surface-induced dissociation: An effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides, *J. Mass Spectrom. 31*, 339–350.
103. Laskin, J. and Futrell, J. H. (2003) Surface-induced dissociation of peptide ions: kinetics and dynamics, *J. Am. Soc. Mass Spectrom. 14*, 1340–1347.
104. Laskin, J. and Futrell, J. H. (2003) Energy transfer in collisions of peptide ions with surfaces, *J. Chem. Phys. 119*, 3413–3420.
105. Little, D. P., Speir, J. P., Senko, M. W., O'Connor, P. B. and McLafferty, F. W. (1994) Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing., *Anal. Chem. 66*, 2809–2815.
106. Fung, Y. M. E., Kjeldsen, F., Silivra, O. A., Chan, T. W. D. and Zubarev, R. A. (2005) Facile disulfide bond cleavage in gaseous peptide and protein cations by ultraviolet photodissociation at 157 nm, *Angew. Chem. Int. Ed. Engl. 44*, 6399–6403.
107. Cui, W. D., Thompson, M. S. and Reilly, J. P. (2005) Pathways of peptide ion fragmentation induced by vacuum ultraviolet light, *J. Am. Soc. Mass Spectrom. 16*, 1384–1398.
108. Madsen, J. A., Boutz, D. R. and Brodbelt, J. S. Ultrafast ultraviolet photodissociation at 193 nm and its applicability to proteomic workflows, *J. Proteome Res. 9*, 4205–4214.
109. Brodbelt, J. (2011) Shedding light on the frontier of photodissociation, *J. Am. Soc. Mass Spectrom. 22*, 197–206.
110. Parthasarathi, R., He, Y., Reilly, J. P. and Raghavachari, K. (2010) New insights into the vacuum UV photodissociation of peptides, *J. Am. Chem. Soc. 132*, 1606–1610.
111. Zubarev, R. A., Kelleher, N. L. and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations., *J. Am. Chem. Soc. 120*, 3265–3266.
112. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, *Proc. Natl. Acad. Sci. USA. 101*, 9528–9533.
113. Zubarev, R. A. (2004) Electron-capture dissociation tandem mass spectrometry, *Curr. Opin. Biotechnol. 15*, 12–16.
114. Zubarev, R. A. (2003) Reactions of polypeptide ions with electrons in the gas phase, *Mass Spectrom. Rev. 22*, 57–77.
115. Savitski, M. M., Kjeldsen, F., Nielsen, M. L. and Zubarev, R. A. (2006) Complementary sequence preferences of electron-capture dissociation and vibrational excitation in fragmentation of polypeptide polycations, *Angew. Chem. Int. Ed. 45*, 5301–5303.
116. Wells, J. M., Chrisman, P. A. and McLuckey, S. A. (2002) "Dueling" ESI: instrumentation to study ion/ion reactions of electrospray-generated cations and anions, *J. Am. Soc. Mass Spectrom. 13*, 614–622.
117. Tarantin, N. I. (1999) Methods of measuring masses in nuclear physics. The basis of mass analysis is the dispersion of ions or charged particles, *Phys. Part. Nucl. 30*, 167–194.
118. Wollnik, H. (1999) Ion optics in mass spectrometers, *J. Mass Spectrom. 34*, 991–1006.

119. Dempster, A. J. (1918) a new method of positive ray analysis, *Phys. Rev. 11*, 316–325.
120. Mattauch, J. (1936) A Double-Focusing Mass Spectrograph and the Masses of N15 and O18, *Phys. Rev. 50*, 617–623.
121. Leverence, R., Mason, A. B. and Kaltashov, I. A. (2010) Noncanonical interactions between serum transferrin and transferrin receptor evaluated with electrospray ionization mass spectrometry, *Proc. Natl. Acad. Sci. USA. 107*, 8123–8128.
122. Stephens, W. E. (1946) A pulsed mass spectrometer with time dispersion, *Phys. Rev. 69*, 691–691.
123. Cameron, A. E. and Eggers, D. F. (1948) An ion velocitron, *Rev. Sci. Instrum. 19*, 605–607.
124. Mamyrin, B. A. (2001) Time-of-flight mass spectrometry (concepts, achievements and prospects), *Int. J. Mass Spectrom. 206*, 251–266.
125. Cotter, R. J. (1997) *Time-of-flight mass spectrometry: instrumentation and applications in biological research*, American Chemical Society, Washington, DC.
126. Cordero, M. M., Cornish, T. J., Cotter, R. J. and Lys, I. A. (1995) Sequencing peptides without scanning the reflectron—post-source decay with a curved-field reflectron time-of-flight mass-spectrometer, *Rapid Commun. Mass Spectrom. 9*, 1356–1361.
127. Uphoff, A. and Grotemeyer, J. (2003) The secrets of time-of-flight mass spectrometry revealed, *Eur. J. Mass Spectrom. 9*, 151–164.
128. Cordero, M. M., Cornish, T. J., Cotter, R. J. and Lys, I. A. (1995) Sequencing peptides without scanning the reflectron - post-source decay with a curved-field reflectron time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom. 9*, 1356–1361.
129. Cornish, T. J. and Cotter, R. J. (1993) Tandem time-of-flight mass spectrometer, *Anal. Chem. 65*, 1043–1047.
130. Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L. and Burlingame, A. L. (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer, *Anal. Chem. 72*, 552–558.
131. Beussman, D. J., Vlasak, P. R., Mclane, R. D., Seeterlin, M. A. and Enke, C. G. (1995) Tandem reflectron time-of-flight mass spectrometer utilizing photodissociation, *Anal. Chem. 67*, 3952–3957.
132. Paul, W. and Steinwedel, H. (1953) Ein Neues Massenspektrometer Ohne Magnetfeld, *Z. Naturforsch. A 8*, 448–450.
133. Paul, W. and Raether, M. (1955) Das Elektrische Massenfilter, *Z. Phys. 140*, 262–273.
134. Paul, W., Reinhard, H. P. and Vonzahn, U. (1958) Das Elektrische Massenfilter Als Massenspektrometer Und Isotopentrenner, *Z. Phys. 152*, 143–182.
135. Leary, J. J. and Schmidt, R. L. (1996) Quadrupole mass spectrometers: An intuitive leak at the math, *J. Chem. Educ. 73*, 1142–1145.
136. Steel, C. and Henchman, M. (1998) Understanding the quadrupole mass filter through computer simulation, *J. Chem. Educ. 75*, 1049–1054.
137. Paul, W. (1990) Electromagnetic traps for charged and neutral particles, *Rev. Mod. Phys. 62*, 531–540.
138. March, R. E. (1997) An introduction to quadrupole ion trap mass spectrometry, *J. Mass Spectrom. 32*, 351–369.
139. Jonscher, K. R. and Yates, J. R., 3rd. (1997) The quadrupole ion trap mass spectrometer - a small solution to a big challenge, *Anal. Biochem. 244*, 1–15.
140. Rocher, F., Favre, A., Gonnet, F. and Tabet, J. C. (1998) Study of ghost peaks resulting from space charge and non-linear fields in an ion trap mass spectrometer, *J. Mass Spectrom. 33*, 921–935.
141. Favre, A., Gonnet, F. and Tabet, J. C. (2001) Perturbation of ion trajectories by resonant excitation leads to occurrence of ghost peaks, *Rapid Commun. Mass Spectrom. 15*, 446–450.
142. March, R. E. (1998) Quadrupole ion trap mass spectrometry: Theory, simulation, recent developments and applications, *Rapid Commun. Mass Spectrom. 12*, 1543–1554.
143. March, R. E. (2000) Quadrupole ion trap mass spectrometry: a view at the turn of the century, *Int. J. Mass Spectrom. 200*, 285–312.
144. Drees, T. and Paul, W. (1964) Beschleunigung von elektronen in einem plasmabetatron, *Z. Chem. 180*, 340–361.
145. Hager, J. (2002) A new linear ion trap mass spectrometer, *Rapid Commun. Mass Spectrom. 16*, 512–526.
146. Schwartz, J. C., Senko, M. W. and Syka, J. E. (2002) A two-dimensional quadrupole ion trap mass spectrometer, *J. Am. Soc. Mass Spectrom. 13*, 659–669.
147. Hager, J. W. and Le Blanc, J. C. Y. (2003) Product ion scanning using a Q-q-Q(linear ion trap) (Q TRAP (TM)) mass spectrometer, *Rapid Commun. Mass Spectrom. 17*, 1056–1064.
148. Goudsmit, S. A. (1948) A time-of-flight mass spectrometer, *Phys. Rev. 74*, 622–623.
149. Hays, E. E., Richards, P. I. and Goudsmit, S. A. (1951) Mass measurements with a magnetic time-of-flight mass spectrometer, *Phys. Rev. 84*, 824–829.
150. Smith, L. G. (1951) A new magnetic period mass spectrometer, *Rev. Sci. Instrum. 22*, 115–116.
151. Smith, L. G. (1952) Recent developments with the mass synchrometer, *Phys. Rev. 85*, 767–767.
152. Sommer, H. and Thomas, H. A. (1950) Detection of magnetic resonance by ion resonance absorption, *Phys. Rev. 78*, 806–806.
153. Sommer, H., Thomas, H. A. and Hipple, J. A. (1951) The Measurement of e/m by cyclotron resonance, *Phys. Rev. 82*, 697–702.
154. Ionov, N. I., Mamyrin, B. A. and Fiks, V. B. (1953) Resonance magnetic mass spectrometer of high resolving power, *Z. Tekhn. Fiz. 23*, 2104–2106.
155. Wobschall, D. (1965) Ion cyclotron resonance spectrometer, *Rev. Sci. Instrum 36*, 466–475.

156. Comisarow, M. B. and Marshall, A. G. (1974) Fourier transform ion cyclotron resonance spectroscopy, *Chem. Phys. Lett. 25*, 282–283.

157. Marshall, A. G. (2000) Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development, *Int. J. Mass Spectrom. 200*, 331–356.

158. Marshall, A. G. and Hendrickson, C. L. (2002) Fourier transform ion cyclotron resonance detection: principles and experimental configurations, *Int. J. Mass Spectrom. 215*, 59–75.

159. Taylor, P. K. and Amster, I. J. (2003) Space charge effects on mass accuracy for multiply charged ions in ESI FTICR, *Int. J. Mass Spectrom. 222*, 351–361.

160. Williams, E. R., Henry, K. D. and Mclafferty, F. W. (1990) Multiple remeasurement of ions in Fourier transform mass spectrometry, *J. Am. Chem. Soc. 112*, 6157–6162.

161. Speir, J. P., Gorman, G. S., Pitsenberger, C. C., Turner, C. A., Wang, P. P. and Amster, I. J. (1993) Remeasurement of ions using quadrupolar excitation Fourier transform ion cyclotron resonance spectrometry, *Anal. Chem. 65*, 1746–1752.

162. Bruce, J. E., Cheng, X., Bakhtiar, R., Wu, Q., Hofstadler, S. A. anderson, G. A. and Smith, R. D. (1994) Trapping, detection and mass measurement of individual ions in a Fourier transform ion cyclotron resonance mass spectrometer, *J. Am. Chem. Soc. 116*, 7839–7847.

163. Sannes-Lowery, K. A. and Hofstadler, S. A. (2000) Characterization of multipole storage assisted dissociation: implications for electrospray ionization mass spectrometry characterization of biomolecules., *J. Am. Soc. Mass Spectrom. 11*, 1–9.

164. Horn, D. M., Ge, Y. and McLafferty, F. W. (2000) Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins, *Anal. Chem. 72*, 4778–4784.

165. Amster, I. J. (1996) Fourier transform mass spectrometry, *J. Mass Spectrom. 31*, 1325–1337.

166. Marshall, A. G., Hendrickson, C. L. and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer, *Mass Spectrom. Rev. 17*, 1–35.

167. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis, *Anal. Chem. 72*, 1156–1162.

168. Hardman, M. and Makarov, A. A. (2003) Interfacing the Orbitrap mass analyzer to an electrospray ion source, *Anal. Chem. 75*, 1699–1705.

169. Kingdon, K. H. (1923) A method for the neutralization of electron space charge by positive ionization at very low gas pressures., *Phys. Rev. 21*, 408–418.

170. Thevis, M., Makarov, A. A., Horning, S. and Schänzer, W. (2005) Mass spectrometry of stanozolol and its analogues using electrospray ionization and collision-induced dissociation with quadrupole-linear ion trap and linear ion trap-Orbitrap hybrid mass analyzers, *Rapid Commun. Mass Spectrom. 19*, 3369–3378.

171. Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M. and Cooks, R. G. (2005) The Orbitrap: a new mass spectrometer, *J. Mass Spectrom. 40*, 430–443.

172. McDaniel, E. W., Martin, D. W. and Barnes, W. S. (1962) Drift-tube mass spectrometer for studies of low-energy ion-molecule reactions, *Rev. Sci. Instrum. 33*, 2–7.

173. Kebarle, P. and Hogg, A. M. (1965) Mass-spectrometric study of ions at near atmospheric pressures. I. Ionic polymerization of ethylene, *J. Chem. Phys. 42*, 668–674.

174. Clemmer, D. E. and Jarrold, M. F. (1997) Ion mobility measurements and their applications to clusters and biomolecules, *J. Mass Spectrom. 32*, 577–592.

175. Hoaglund, C. S., Valentine, S. J., Sporleder, C. R., Reilly, J. P. and Clemmer, D. E. (1998) Three-dimensional ion mobility TOFMS analysis of electrosprayed biomolecules, *Anal. Chem. 70*, 2236–2242.

176. Hudgins, R. R., Woenckhaus, J. and Jarrold, M. F. (1997) High resolution ion mobility measurements for gas phase proteins: correlation between solution phase and gas phase conformations, *International Journal of Mass Spectrometry 165*, 497–507.

177. Wyttenbach, T., Kemper, P. R. and Bowers, M. T. (2001) Design of a new electrospray ion mobility mass spectrometer, *International Journal of Mass Spectrometry 212*, 13–23.

178. Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. and Scrivens, J. H. (2007) An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument, *Int. J. of Mass Spectrom. 261*, 1–12.

179. Bernstein, S. L., Liu, D. F., Wyttenbach, T., Bowers, M. T., Lee, J. C., Gray, H. B. and Winkler, J. R. (2004) Alpha-synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation, *J. Am. Soc. Mass Spectrom. 15*, 1435–1443.

180. Borysik, A. J. H., Read, P., Little, D. R., Bateman, R. H., Radford, S. E. and Ashcroft, A. E. (2004) Separation of beta (2)-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry, *Rapid Commun. Mass Spectrom. 18*, 2229–2234.

181. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water, *Science 310*, 1658–1661.

182. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. and Heck, A. J. R. (2010) Ion mobility mass spectrometry of proteins and protein assemblies, *Chem. Soc. Rev. 39*, 1633–1655.

183. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. and Hill, H. H. (2008) Ion mobility-mass spectrometry, *J. Mass Spectrom. 43*, 1–22.

184. Chernushevich, I. V., Ens, W. and Standing, K. G. (1999) Orthogonal injection TOF MS for analyzing biomolecules, *Anal. Chem. 71*, 452A–461A.

185. Doroshenko, V. M. and Cotter, R. J. (1998) A quadrupole ion trap time-of-flight mass spectrometer with a parabolic reflectron, *J. Mass Spectrom. 33*, 305–318.

186. Parks, B. A., Jiang, L., Thomas, P. M., Wenger, C. D., Roth, M. J., Boyne, M. T., Burke, P. V., Kwast, K. E. and

Kelleher, N. L. (2007) Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers, *Anal. Chem. 79*, 7984–7991.

187. Kaplan, D. A., Hartmer, R., Speir, J. P., Stoermer, C., Gumerov, D., Easterling, M. L., Brekenfeld, A., Kim, T., Laukien, F. and Park, M. A. (2008) Electron transfer dissociation in the hexapole collision cell of a hybrid quadrupole-hexapole Fourier transform ion cyclotron resonance mass spectrometer, *Rapid Commun. Mass Spectrom. 22*, 271–278.
188. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K. and Horning, S. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer, *Anal. Chem. 78*, 2113–2120.
189. Makarov, A., Denisov, E., Lange, O. and Horning, S. (2006) Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer, *J. Am. Soc. Mass Spectrom. 17*, 977–982.

4

MASS SPECTROMETRY BASED APPROACHES TO STUDY BIOMOLECULAR HIGHER ORDER STRUCTURE

Obtaining higher order structures of biopolymers is usually a first step in analyzing their behavior and understanding function. While X-ray crystallography and high-field NMR undoubtedly provide the highest quality information, limitations of these techniques (as discussed in Chapter 2) often make such high-resolution structures unavailable for a variety of important proteins and their assemblies. Other biophysical techniques, (e.g., SEC, and scattering methods) often provide information that can be used for evaluating quaternary structure of protein assemblies; however, the resolution is typically low. This chapter discusses various MS based approaches to evaluate higher order structure at various levels of spatial resolution when the crystallographic and NMR data are unavailable or insufficient. This chapter begins with a presentation of methods used to determine the composition and stoichiometry of protein assemblies and other noncovalent complexes of biopolymers, the domain of "native" ESI MS. We will also briefly review additional information that can be deduced from native ESI MS measurements. This includes understanding the hierarchical organization of protein quaternary structure (based on controlled fragmentation of protein complexes in the gas phase) and estimating solvent-accessible surface areas in proteins and protein complexes (based on the number of charges accommodated by the protein). We will then proceed to review MS-based methods used to probe biomolecular topology and topography. These methods generally utilize chemical cross-linking to characterize tertiary and quaternary contacts by generating proximity maps. Finally, we will discuss various methods of mapping solvent accessibility of protein segments. In addition to permanent covalent modification of protein surfaces in both specific and nonspecific fashion, we will briefly discuss application of hydrogen–deuterium exchange to mapping protein–protein interfaces.

4.1. DIRECT METHODS OF STRUCTURE CHARACTERIZATION: NATIVE ELECTROSPRAY IONIZATION MASS SPECTROMETRY

4.1.1. Preservation of Noncovalent Complexes in the Gas Phase: Stoichiometry of Biomolecular Assemblies

It was not too long ago that probing the quaternary structure of proteins in solution using MS was met with skepticism in many quarters of the biophysical community. Indeed, water is essential to biological processes in general and to the protein structure and function in particular (1). Significant alterations of the protein structure upon dramatic changes in its environment (i.e., complete removal of solvent) are expected and in fact have been documented in many cases (2,3). Protein behavior in a solvent-free environment is now being actively looked at, as such studies hold the promise of providing a distinction between the "intrinsic" and "externally imposed" properties of biological macromolecules (4). One general conclusion that is particularly important within the context of this chapter relates to the fate of protein assemblies upon desorption from solution to the gas phase. The existing body of knowledge suggests

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

that most types of intra- and intermolecular interactions in proteins are preserved *in vacuo* (5,6), with the notable exception of hydrophobic interactions [although a recent report suggests that macromolecular associations held together exclusively by hydrophobic interactions may also survive, especially if there is a high steric complementarity between the binding partners (7)].

Numerous studies have demonstrated that noncovalent macromolecular complexes can survive the transition from solution to the gas phase when using relatively mild desolvation conditions in the ESI MS interface. The two parameters that are usually most critical for the survival of noncovalent complexes upon this transition are the ESI interface temperature and the electrical field in the ion desolvation region, which determines the average kinetic energy of ions undergoing frequent collisions with neutral molecules in this region. Keeping these parameters at relatively low levels allows the composition of macromolecular assemblies to be determined reliably and with minimal sample consumption (Fig. 4.1). When executed carefully, such experiments produce correct (i.e., confirmed by other methods) information on the stoichiometry of multiprotein complexes (8–12) and provide information on smaller ligands (metal ions, small organic molecules, etc.) that are present in these multiunit assemblies. While selecting mild conditions in the ESI MS interface region is critical for success of such experiments, another critical aspect in experimental design is selection of a solvent system that is compatible with ESI (see Section 3.2.2 for more detail), and yet has the physicochemical parameters (most importantly pH and ionic strength) that are as close as possible to that of the *native* environment of the proteins [hence the frequently used name of this technique, native MS (11)]. This is usually achieved by using aqueous solutions of volatile salts (e.g., ammonium acetate or ammonium bicarbonate) at appropriate concentration levels, which produce pH close to neutral.

Determination of the number of polypeptide subunits in relatively small homooligomers usually does not require very high accuracy in mass measurements. However, a serious problem may arise if the complex is composed of a number of different polypeptide chains or else contains low molecular weight components (metal ions, organic cofactors, etc.), which need to be accounted for. As mentioned above, preservation of noncovalent interactions is often achieved by reducing the efficiency of the ion desolvation step, and the survival of larger macromolecular complexes critically depends on the presence of a significant number of residual solvent molecules and small counterions (13,14). Therefore, preservation of noncovalent complexes in the gas phase is typically done at the expense of mass accuracy, which is critical for correct assignment of the small molecular weight components of a high molecular weight complex.

In addition to adduct formation, which increases the apparent mass of the macromolecular species, the mass

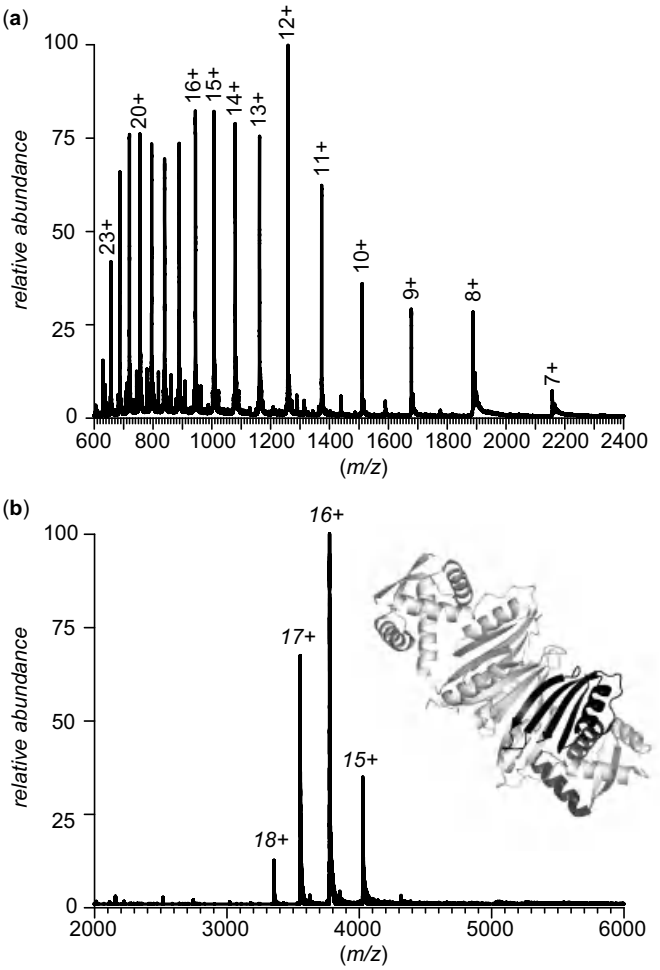


Figure 4.1. The ESI mass spectra of bacterial enzyme NikR acquired under denaturing (*a*) and near-native conditions (*b*). Protein dissociation upon denaturation allows the mass of the monomeric unit to be measured with high precision ($15{,}093 \pm 1$ Da). Mass measurements carried out under near-native conditions ($60{,}372 \pm 2$ Da) indicate that the protein exists in solution exclusively in the tetrameric form, consistent with the crystal structure (shown in panel *b*; PDB accession number 1Q5V).

measurement process is frequently complicated by the ambiguity of charge-state assignment of ions in ESI mass spectra. The task of correct charge-state assignment becomes increasingly difficult for larger macromolecular systems, and is greatly exacerbated by the heterogeneity of ionic species. Several methods have been introduced in recent years that deal with this problem, most of which rely upon finding the best fit for the experimental data using various computational algorithms (9,15,16). The basic assumption of these approaches is that the mass of the noncovalent complex (or any macromolecular ionic species) is constant and allows the correct charge states in a peak series to be determined using an optimization procedure. The correct charge-state assignment is the one that yields

the minimal standard deviation S for the entire series of ionic species (17):

$$S = \sqrt{\sum_{i=1}^{N-1} (k_i \cdot z_i - \langle m \rangle)^2} \qquad (4\text{-}1\text{-}1)$$

where k_i is the apparent *m/z* for the *i*th ionic peak in the data set containing N peaks representing the macromolecule; z_i is the charge assigned to this peak, and $\langle m \rangle$ is the mass of the macromolecule (defined as the mean mass averaged across the entire set of N ionic species. Sophisticated algorithms are capable of defining the boundaries of mixed ionic signals when multiple species are present in solution and do not have limitations vis-à-vis *m/z* windows (17).

4.1.2. Utilization of Ion Chemistry in the Gas Phase to Aid Interpretation of ESI MS Data

The optimization methods discussed in the preceding paragraph have been utilized successfully to interpret ESI MS data for very large (>1 MDa), but relatively homogeneous, proteins and protein complexes. However, they frequently fail when applied to highly heterogeneous systems. The structural heterogeneity of glycoproteins with high carbohydrate content and other polydisperse macromolecules poses a serious challenge to ESI MS analysis for two reasons. First, the population of ionic species with different masses distributed over a relatively narrow *m/z* range often gives rise to unresolved or poorly resolved ionic signals and may lead to significant overlap of ion peaks representing different charge states. Second, as we will see in Section 4.1.4, the extent of multiple charging in ESI MS depends on the physical size of the macromolecules. In the case of polydisperse systems, this may result in a situation where the measured average ionic mass (or mass distribution) depends on its charge, which obviously complicates deconvolution of ESI MS data (18). A recently introduced approach to deal with this problem combines complexity reduction (mass-selection of a narrow distribution of ionic species from a heterogeneous mixture) and gas-phase ion chemistry (electron transfer reactions) to induce partial reduction of the ionic charge. The resulting spectra are devoid of complexity and are easy to interpret, allowing correct mass assignment to be made (Fig. 4.2).

Collisionally activated dissociation (CAD) is another type of ion manipulation in the gas phase that is frequently employed to aid interpretation of ESI mass spectra of noncovalent complexes. We already mentioned that preservation of noncovalent interactions is often done by reducing the efficiency of the ion desolvation step, which inevitably promotes generation of ions with extensive adduct binding. To improve the accuracy of mass measurements under these circumstances, Amster and co-workers (19) suggested to

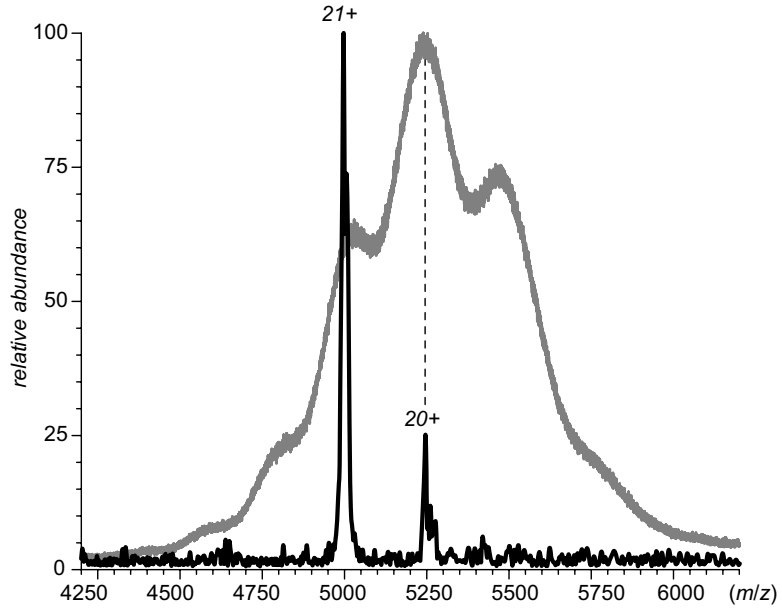


Figure 4.2. The ESI mass spectra of α-galactosidase (3 μM in 50 mM ammonium acetate) acquired with a hybrid quadrupole–time of flight mass spectrometer (TOF MS, gray trace) and the results of incomplete reduction of a narrow population of protein ions mass selected from one of the charge states (black trace). [Adapted with permission from (18). Copyright © 2010 American Chemical Society.]

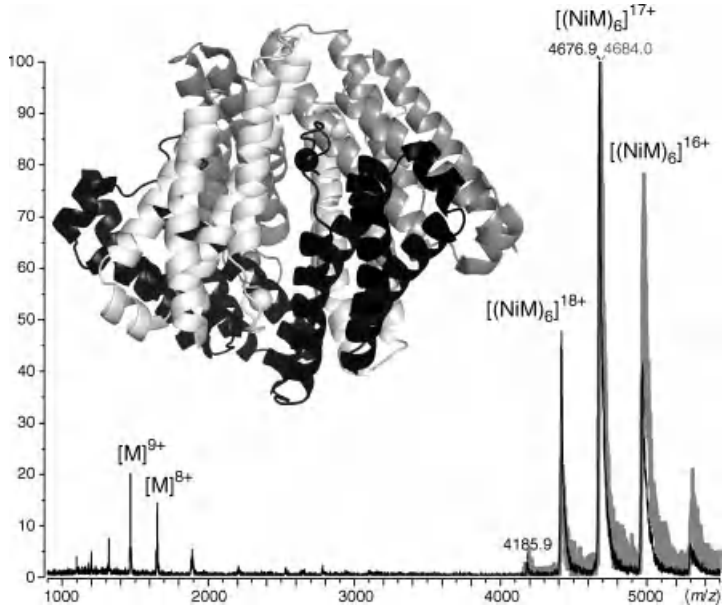


Figure 4.3. The ESI mass spectra of *Streptomyces* nickel superoxide dismutase (NiSOD) (0.125 mg/mL in 5-mM ammonium bicarbonate) acquired under relatively mild conditions in the ESI interface (gray trace) and following collisional activation.

supplement mild ESI MS measurements with those carried out under harsher conditions. Although the latter step leads to partial dissociation of noncovalent complexes in the gas phase, the surviving assemblies appear to have lower residual solvation, and a stepwise increase of the electrostatic field in the interface region eventually results in dissociation of cofactors from the subunits, thereby enabling a reliable identification and an exact count of the low molecular weight species present in each subunit (19).

An example of using limited collisional activation of noncovalently bonded protein complexes in the gas phase to enhance the quality of their mass measurement is shown in Figure 4.3, which presents the work carried out in the authors' laboratory in 1999. An ESI mass spectrum of a metalloenzyme NiSOD from *Streptomyces* (20) acquired under relatively gentle conditions in the ESI interface region revealed the presence of a species whose mass was determined to be 79,594 Da (gray trace in Fig. 4.3). Although this mass is consistent with the notion of NiSOD being a hexamer (average mass of a monomeric unit calculated based on the protein sequence is 13,188.1 Da), it did not allow the exact number of Ni atoms to be determined, as the measured mass exceeded that calculated for the hexamer (where Ni^{2+} is bound to each monomeric unit) by 0.16%. While this might be considered an excellent precision by someone accustomed to polyacrylamide gel electrophoresis (PAGE)- or SEC-based mass measurements, the absolute value of the resulting discrepancy (125 Da) exceeds the mass of two additional nickel atoms. Furthermore, prior to these MS measurements, NiSOD was believed to be a tetramer based on the results of earlier cross-linking measurements, where four cross-linked units were observed as the highest oligomer (20). This contributed further uncertainty toward interpretation and indeed acceptance of the MS data. However, collisional activation of the NiSOD complex in the gas phase (black trace in Fig. 4.3) allowed the mass of the surviving assembly to be measured with much higher precision, which yielded a discrepancy of only 4 Da (0.005%) between the measured mass and the sum of six polypeptide chains and six nickel atoms. It is ironic that the dispute between native ESI MS measurements and cross-linking experiments regarding the correct stoichiometry of NiSOD was settled a couple of years later in favor of ESI MS by the authors of the initial cross-linking work, who published a crystal structure of this enzyme, which was indeed hexameric (21). It is difficult to think of a more gratifying way to have once controversial conclusions proven correct!

Stepwise Dissociation of Macromolecular Complexes in the Gas Phase. Dissociation can also be very effective when analyzing the composition of complex heterooligomeric assemblies containing a significant number of incongruent biopolymer subunits (22–25). In some cases, the order of sequential departure of subunits from the assembly may provide low-resolution information on protein quaternary structure by revealing the identity of interacting subunits in the complex. Even in a situation when all subunits are identical, controlled dissociation experiments can be quite useful, as they can reveal details of the "super-quaternary" structure, that is, hierarchical organization of protein complexes (26,27).

4.1.3. Dissociation of Noncovalent Complexes in the Gas Phase: Can It Lead to Wrong Conclusions?

Despite the great analytical utility of studying the gas-phase dissociation of noncovalent complexes, these processes frequently cause concern. Indeed, it could be argued that the occurrence of uncontrolled fragmentation in the gas phase may lead to incorrect conclusions vis-à-vis composition of the noncovalent assemblies in solution. In most cases, however, these concerns are unfounded, as the common mechanism of dissociation of noncovalent complexes in the gas phase [*asymmetric charge partitioning* (28)] gives rise to a distinct dissociation pattern that is very different from the disassembly processes occurring in solution. Asymmetric charge partitioning proceeds through ejection of a single polypeptide chain from a metastable complex, with the ejected unit carrying a disproportionately high number of charges (29). An example of such a process is shown in Figure 4.4, where mild collisional activation of large protein complexes (the 14-mer molecular chaperone GroEL) in the ESI interface results in ejection of a highly charged monomeric ion from the complex. The remainder of the complex (13-mer) retains a disproportionately low fraction of the initial number of charges (average charge 43+), giving rise to ionic signal at the high end of the *m/z* scale. The anomalous charge density of the products of dissociation is a clear indication that these species are generated in the gas phase rather than in solution (30).

The origin of asymmetric charge partitioning is likely to be the electrostatic repulsion within the ESI generated multicharged noncovalent complexes, in which case the fragmentation process should resemble Coulombic explosion of a liquid droplet charged to the Rayleigh limit (31). The commonly observed asymmetric dissociation in this case would correspond to a *fine fission* mode, which is initiated in critically charged droplets by formation of a Taylor cone, through which ejection of a small mass of liquid carrying a disproportionately large fraction of the total charge occurs (13). It seems plausible that fluctuations of charge density on the surface of a globular noncovalent assembly in the gas phase would lead to formation of a local instability similar to the Taylor cone, followed by ejection of highly charged matter. The charge-density fluctuations are likely to occur within the solvation shell of the noncovalent complex. The solvation shell is composed mostly of the ionic and polar solvent components, which may shuttle the charges across the surface of the complex, eventually

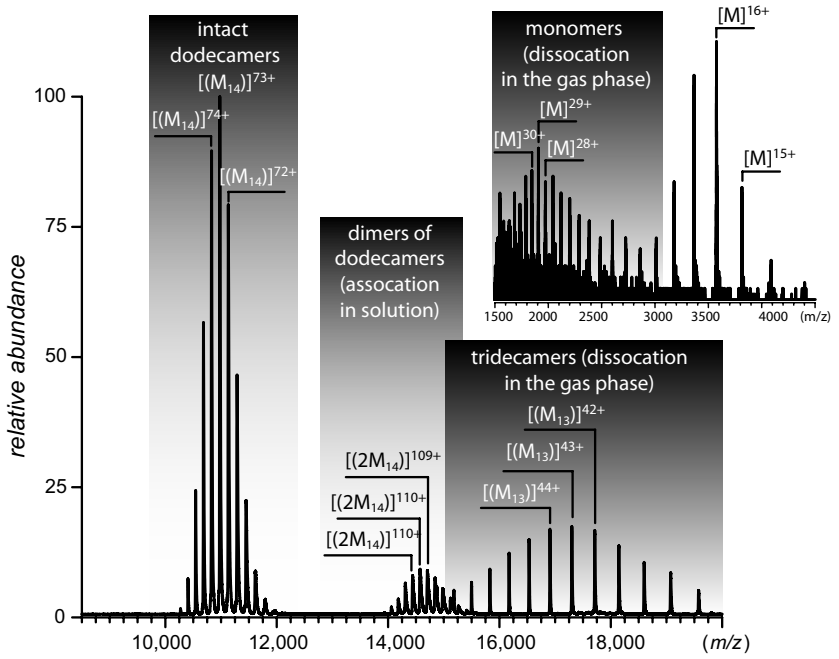


Figure 4.4. The ESI mass spectrum of GroEL acquired under near-native conditions in solution (100-mM ammonium acetate) and mild collisional activation in the ESI interface. The low *m/z* region of the spectrum is shown in the inset. Highly charged monomers and low charge density tridecamers are products of dissociation of tetradecameric structures in the gas phase. Oligomerization of GroEL tetradecamers (formation of $2M_{14}$ species) is likely caused by increased protein concentration in ESI droplets as a result of solvent evaporation. [Adapted with permission from (30). Copyright © 2008 American Society for Mass Spectrometry.]

leading to formation of a local instability (a process that is obviously facilitated by collisional activation of the complex). Since the ejection of soft matter in the fine fission mode occurs through a fine jet, the departing monomer will necessarily unravel in the process of (but not prior to) its ejection. Since the jet is formed by a single subunit, it will not disintegrate, unless the departing monomer contains noncovalently bound ligands and cofactors.

Such behavior can be seen in Figure 4.5, where asymmetric dissociation of a globular tetrameric complex (hemoglobin tetramer $(\alpha^*\beta^*)_2$, where * denotes a heme group attached to a globin chain) results in ejection of a highly charged monomer (either α or β), whose heme group does not necessarily accompany the chain during this event. This results in the appearance of a rather exotic ionic species in the mass spectrum, such as a trimer carrying four heme groups (e.g., $(\alpha^*{}_2\beta^{**})^{8+}$ in the expanded region of Fig. 4.5b). Such extreme behavior exhibited by the protein assembly upon its dissociation in the gas phase is very different from hemoglobin tetramer dissociation in solution, which is known to proceed via formation of heterodimers (i.e., $(\alpha^*\beta^*)_2 \rightarrow 2\ \alpha^*\beta^*$, followed by dimer dissociation to monomers). In general, dissociation of protein assemblies in the gas phase almost always manifests itself in ESI mass spectra by giving rise to ionic species of either exotic composition or anomalous charge density. Therefore, in most cases it is not too difficult to make a clear distinction between the dissociation processes occurring in solution and in the gas phase. In addition the experimental conditions can be adjusted to minimize or completely eliminate the former.

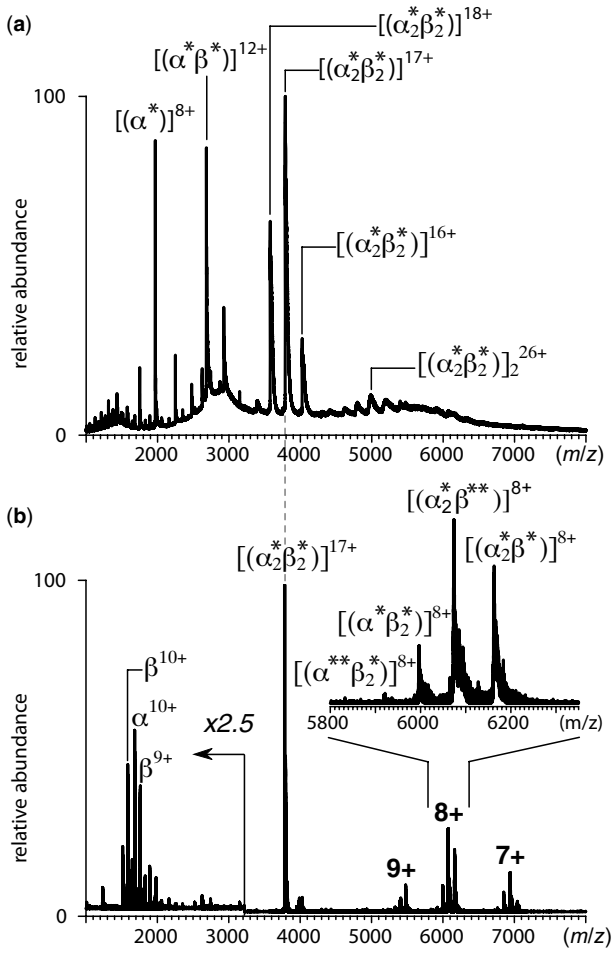


Figure 4.5. The ESI mass spectra of sickle cell human hemoglobin HbS under near-native conditions in solution. Panel *a* shows an HbS spectrum acquired under gentle conditions in the ESI interface that preserve weak noncovalent interaction. Panel *b* shows a CAD mass spectrum of HbS tetramer (charge state 17+ isolated as a precursor ion). Expanded region on panel *b* shows a detailed view of the trimeric region of the fragment ions mass spectrum. The asterisk represents the bond hence cofactor. [Adapted with permission from (13). Copyright © 2006 Elsevier.]

4.1.4. Evaluation of Macromolecular Shape in Solution: The Extent of Multiple Charging in ESI MS

Characterization of multiunit protein assemblies discussed in the previous sections relies on measuring the masses of protein ions and their complexes. However, the protein ion charge can also provide useful information. In fact, it was not long after the acceptance of ESI MS as a tool for measuring masses of macromolecular ions that observations were made linking the extent of multiple charging to protein geometry in solution (32,33). The relationship between the charge on a protein ion and its size (geometry) has been the focus of numerous investigations. One particularly useful tool that has emerged from these studies is the analysis of conformational dynamics in solution, which allows onset of protein unfolding to be detected by monitoring protein ion charge-state distributions in ESI MS (considered in more detail in Chapter 5). However, even in the absence of unfolding events, analysis of charges accommodated by protein ions may provide useful information regarding their higher order structure.

The average charge of a protein ion tends to increase as the protein size (mass) increases. This is particularly obvious under denaturing conditions, as the *m/z* values of most proteins tend to fall within the same range (800–2500 u) irrespective of their masses. The average charges of protein ions generated under near-native conditions also increase as the protein mass increases, although not as fast, giving rise to a progressive increase in the *m/z* values (Fig. 4.6). The mass spectrum shown in Figure 4.6 is noteworthy for another reason, as it clearly demonstrates that, unlike the mass, the charge of protein complex ions does not follow the principle of additivity, that is the average charge of the protein complex ion is significantly less than the sum of the

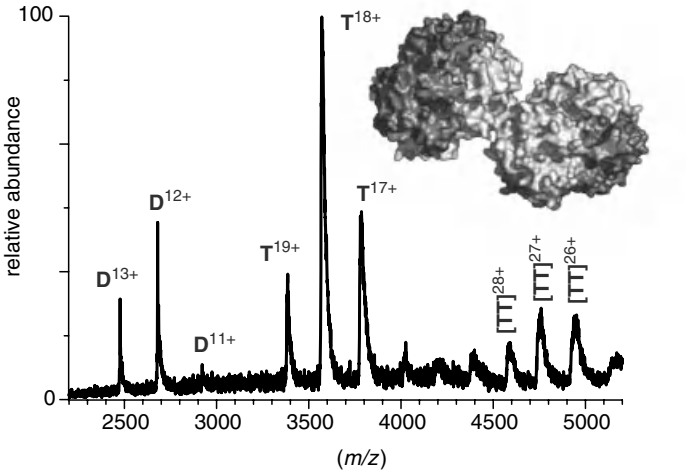


Figure 4.6. The ESI mass spectrum of human sickle cell hemoglobin (40 μM solution, calculation of molar concentration is based on the tetrameric structure) acquired at neutral pH in 10 mM ammonium acetate. The labeled ionic species represent dimers (D), tetramers (T), and octamers (TT). [Adapted with permission from (38). Copyright © 2005 American Chemical Society.]

charges of ions representing the individual constituents of the complex.

The average number of charges accommodated by protein ions generated under native conditions in solution was shown by Fenselau et al. (34) to exhibit a strong dependence on the surface area of the native conformation in solution. de la Mora (35), Heck and van Den Heuvel (36), and Nesatyy and Suter (37) analyzed compiled sets of ESI MS data for a variety of globular proteins under native conditions and used Dole's Charged Residue Model (CRM, see Appendix) to rationalize the apparent correlation between the calculated radius of gyration of the proteins in solution and their average and highest charges in the gas phase, as determined from the ESI mass spectra. While these analyses clearly showed a general trend, the observed correlations were not perfect. This is not particularly surprising, since the experimental conditions used to acquire the individual data sets were very different. As a result, the outcomes of such measurements were affected by extrinsic factors to various degrees, causing random deviations from a putative protein ion charge–protein geometry correlation. To eliminate these deviations, we used a set of proteins, ranging from the small polypeptide insulin (5 kDa) to a large multiunit protein ferritin (500 kDa), to acquire MS data under identical conditions aiming to find a correlation between the average ionic charge and parameters of protein geometry in solution, such as surface area (38). All measurements were carried out under conditions that minimize variation of the extrinsic factors affecting protein ion charge state distributions (38,39). The experimentally determined empirical correlation ($N = N_o + S^{\alpha}$) is presented in Figure 4.7; as can be seen from the discussion below, this correlation is in excellent agreement with the current view of ESI processes and the predictions of the CRM.

According to the CRM model, generation of protein ions is caused by an efficient droplet atomization process, which is a consequence of the *Rayleigh instability*. The instability criterion is given by (40):

$$\gamma = \frac{(Ne)^2}{8\pi\varepsilon_o d_o^3} \qquad (4\text{-}1\text{-}2)$$

where γ is the solvent surface tension, ε_o is the permittivity of vacuum, Ne represents the net charge of the droplet, and d_o is its diameter. The droplets generated during disintegration of the jet emitted from the unstable droplet (*progeny droplets*) have a narrow size distribution and are close to the instability criterion, which means they undergo fission soon after separation from the main droplet. In contrast, a second fission of the parent droplet would require substantial solvent evaporation in order to bring it back to instability conditions (Eq. 4-1-2).

The protein molecule is likely to remain in the parent droplet following the Coulombic explosion, if it was residing in its interior just prior to the fission event. At the same time, protein molecules positioned close to the droplet surface are likely to be ejected with the progeny droplets. In any event, the fission process is likely to occur several times prior to generating a droplet whose size is barely adequate for encapsulating the protein molecule. If protein structure disruption does not occur or is minimal during the ESI process, the geometry of the native conformation will actually be one of the major determinants of the physical size of the smallest droplet still capable of encapsulating the protein. This would set a limit for the total charge accumulated by the protein upon complete evaporation of any residual solvent from such a droplet. Similar arguments were used by de la Mora (35) (in addition to assuming a near-spherical shape of the protein molecule in its native conformation), to predict the following limit on the number of charges accumulated by protein ions:

$$N = \frac{8}{e}\left(\pi\gamma\varepsilon_o R_i^3\right)^{1/2} = \frac{4}{e}\left(\frac{\gamma^2\varepsilon_o^2}{4\pi}\cdot S_i^3\right)^{1/4} \qquad (4\text{-}1\text{-}3)$$

where e is the elementary charge, and R_i (S_i) is the radius (surface area) of the globular protein ion in spherical approximation.

A spherical approximation, however, does not provide adequate description of shape for the majority of proteins, even globular ones. In order to extend Eq. 4-1-3 to proteins

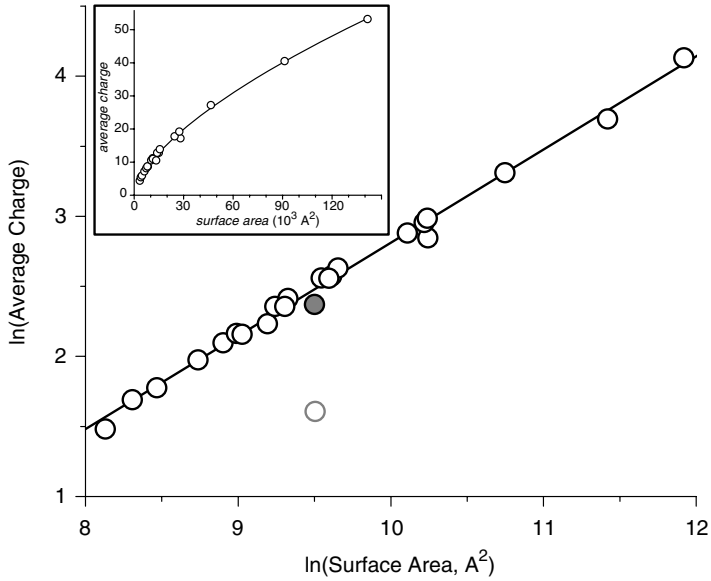


Figure 4.7. Correlation between the average charge state of protein ions generated by ESI under near-native conditions (10 mM ammonium acetate, pH adjusted to 7) and their surface areas in solution, whose calculation was based upon their crystal structures. The data are plotted in the logarithmic scale (a graph plotted in the linear scale is shown in the inset). A gray-shaded dot represents a pepsin data point, not used in calculation of the least-squares linear fit for the entire data set. An open gray circle represents the highest charge of pepsin if the extent of multiple charging were limited by the number of basic residues within the protein molecule. [Adapted with permission from (89). Copyright © 2005 American Chemical Society.]

whose shapes deviate significantly from the ideal sphere, we will consider the electrohydrodynamic processes of parent droplet disintegration and progeny droplet formation in greater detail. Disintegration of the parent droplet in the fine fission mode results in ejection of a jet in the direction of the external field gradient. The jet itself is unstable and eventually disintegrates, giving rise to an ensemble of progeny droplets. A critical question here is how the charge on the jet is distributed among these droplets. There is apparent disagreement in the literature as to what the scaling law is. de Juan and de la Mora (41) suggested that the jet charge "*remains nearly tied to the liquid during the break-up process*," so that the charge density is nearly constant for a given spray:

$$q \sim V \sim R^3 \tag{4-1-4}$$

where V is the volume and R is the radius of a given progeny droplet immediately after jet disintegration. In this case, the charge is assumed to be tied to the bulk volume of the jet. The bulk of the solvent, however, is a conducting liquid, which would remain charge- and field-free in the quasistatic approximation, forcing the net charge to be distributed along the surface of the jet. Therefore, it appears that constancy of the charge–volume density (Eq. 4-1-4) should be substituted with constancy of the charge–surface (S) density:

$$q \sim S \sim R^2 \tag{4-1-5}$$

A more rigorous consideration of these processes by Hartman et al. (42) led to the suggestion that progeny droplet charge–size correlation is even weaker:

$$q \sim V^{1/2} \sim S^{3/4} \sim R^{3/2} \tag{4-1-6}$$

While most progeny droplets assume near-spherical shape following jet disintegration, some internal constraints (e.g., the presence of nonspherical particles inside the jet), may lead to generation of nonspherical droplets as well. In order to estimate the charge transferred to such droplets, one would need to use the surface of the nonspherical particle as a limiting factor:

$$q \sim \left(\frac{S}{a}\right)^{3/4} \tag{4-1-7}$$

where a is a measure of deviation of the droplet surface from that of the encapsulated particle. This correction factor is

expected to be close to unity for particles with topologically simple (smooth) surfaces. However, the value of a is expected to increase as the particle surface topology becomes more complex (due to the presence of cavities, etc.). The scaling law (Eq. 4-1-7) gives the same charge–surface area dependence as the estimation of the charge limit for spherical progeny droplets (Eq. 4-1-3), after taking into account the correction factor a. If the nonideality parameter a does not exhibit significant variation within a set of proteins, one should expect to observe a linear charge–surface dependence in the log–log coordinates with a slope of 0.75. This value is in a remarkably good agreement with the empirically observed correlation shown in Fig. 4.7 (38):

$$\ln(N) = A + \alpha \cdot \ln(S), \quad \alpha = 0.69 \pm 0.02 \qquad (4\text{-}1\text{-}8)$$

Very similar slopes (0.68 ± 0.02 and 0.66 ± 0.03) were obtained in a more recent study based on a smaller set of proteins (39,43) and a recent comprehensive analysis of published data (44). The charge–surface correlation documented in (38) and confirmed in (39,43,44) can be used to provide reasonable estimates of protein surface areas in solution. Although this method presently cannot rival other established techniques as far as measurement precision, it is useful for the characterization of protein assemblies in solution that are not amenable to analysis using traditional biophysical tools due to their transient nature or heterogeneous character, such as early precursors to protein aggregation (39).

4.1.5. Macromolecular Shape in the Gas Phase: Ion Mobility–Mass Spectrometry

With the general acceptance of the fact that the quaternary structure of proteins, as well as other biopolymers, can be preserved in the gas phase, the next logical question is if the same is true for protein conformation. Should the answer to this question be affirmative, then the experimental methods based on ion mobility–mass spectrometry (IM–MS) can be used to characterize the architecture of proteins and their assemblies beyond the composition and surface area estimates that can be provided by ESI MS alone. The results of initial studies are very encouraging, as they suggest that gas-phase mobility of ions representing folded proteins in solution are significantly higher compared to their unfolded counterparts (5), indicating that at least some "memory" of the solution-phase conformation is retained upon protein transition to the gas phase. Applications of IM–MS to characterization of the quaternary structure of protein assemblies have also yielded encouraging results. For example, IM–MS provides clear evidence (45) that the oligomeric intermediates of viral capsid assemblies are less compact compared to globular proteins of the same size (Fig. 4.8). Ion mobility–mass spectrometry was found to be capable of distinguishing the overall subunit architecture of several

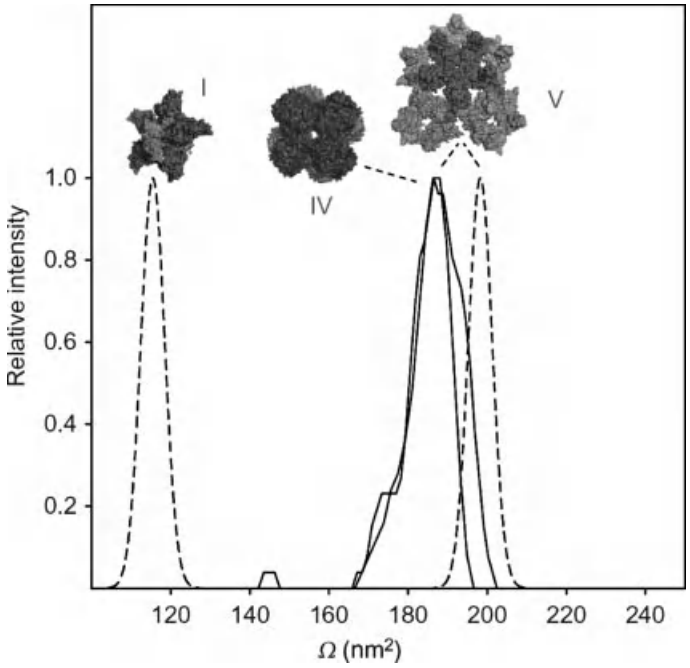


Figure 4.8. The IM–MS data for selected oligomers of human hepatitis B virus (HBV): Overlay of experimental drift time traces converted to collisional cross-section, Ω (solid lines) of the HBV 28-mer (structure V, 447 kDa) and a globular protein complex vanillyl–alcohol oxidase (VAO) 8-mer (IV, 510 kDa), both appearing at m/z10,400. Their Ω values, as well as the peak widths, are similar, even though the VAO 8-mer has a substantially higher mass ($\sim$ 14%). This illustrates that globular proteins of the same mass are generally more compact than the viral intermediates. Also shown are hypothetical Gaussian traces (dashed lines) based on the Ω of the HBV 28-mer modeled from the crystal structure (V) of the virus capsid and a collapsed structure (I). [Adapted from (45).]

protein complexes (46) by showing that their building blocks retain native interactions and do not undergo major rearrangement in the gas phase.

Nevertheless, the solvent-free environment, in many cases, causes dramatic structural alterations in protein conformation. A range of studies indicate that the evolution of protein structure during and after transfer to the gas phase involve side-chain collapse, unfolding, and refolding into new, non-native structures (47). It is not surprising, therefore, that a careful comparison of IM–MS data with the parameters derived from X-ray crystal structures often reveals significant discrepancies (48). These differences are usually consistent with the notion of structural collapse in the gas phase, suggesting that the ESI generated protein assemblies are significantly more compact compared to the corresponding structures in solution (49). The emerging consensus is that great care must be taken when interpreting results of IM–MS measurements, particularly when attempting to correlate IM–MS and crystallography data (48,50). Despite this limitation, IM–MS is likely to continue to enjoy popularity

as a means of assessing higher order protein structure. One of the areas where it may become particularly useful is in developing topological models of protein complexes. This can be done by measuring simultaneously masses and sizes of intact biopolymer assemblies and their subcomplexes in the gas phase, to provide restraints that can be used to filter topological models of protein complexes (46,51).

One of the advantages of using IM–MS based methods for the analysis of large macromolecules and their complexes is the reduction of complexity (due to IM analyzers playing the role of online separation devices). Nevertheless, in the case of large complexes or heterogeneous systems, multiple charging may often lead to difficulties in the interpretation of IM–MS data (52). Loo and co-workers (52) offered an elegant solution to this problem by introducing gas-phase electrophoretic mobility molecular analysis (GEMMA). This technique combines the advantages offered by IM–MS with ionic charge reduction, which greatly decreases the complexity of the ionic population by generating only singly charged molecular species. This technique has been successfully used to obtain electrophoretic mobility diameters (EMD, defined as a diameter of a singly charged sphere with the same electrophoretic mobility as the biopolymer) of a range of large protein complexes (52–54). As is the case for other ion mobility-based measurements, structure collapse in the gas phase appears to affect GEMMA measurements: EMDs for various protein complexes obtained from GEMMA measurements are consistently lower (by $\sim$10%) compared to the parameters derived from their X-ray crystal structures (52) or quasielastic light scattering (55). A more detailed discussion of various applications of IM–MS is beyond the scope of this book, and interested readers are referred to an excellent recent review by Uetrecht et al. (56).

4.1.6. How Relevant Are Native ESI MS Measurements? Restrictions on Solvent Composition in ESI

One aspect of all native ESI MS measurements that must be mentioned in concluding this section is the requirement that the solvent must be compatible with the ESI process. This limits the repertoire of electrolyte systems that can be used to buffer/adjust the ionic strength of protein solutions to several volatile systems (with ammonium acetate and ammonium bicarbonate being the most popular). Although this limitation was considered as very unforgiving in the past, significant improvements in ESI interface design, and the emergence of the nano-ESI technique in particular, provide reasonable grounds for optimism. Higher tolerance of nano-ESI (compared to ESI) for salt and nonvolatile buffers was noted some time ago, at least for small peptides (57). More recently, Williams and co-workers (58) observed that the protein ion signal produced by ESI MS from solutions containing nonvolatile electrolytes can be improved by adding large quantities of ammonium acetate (58), although this may require the ionic strength of the protein solution to be increased several-fold above physiologically relevant levels, which may have negative impact on protein stability (59).

The presence of concentrated NaCl in protein solutions inevitably leads to deterioration of spectral quality in nano-ESI MS, as a result of formation of abundant $Na^+(NaCl)_n$ clusters in the low *m/z* region and protein–sodium chloride adducts at higher *m/z* (Fig. 4.9). Their presence in the mass spectra makes even ion peak assignment a very challenging task, and also is likely to compromise the ability of direct ESI MS approaches to analyze heterogeneous mixtures, since it becomes difficult to distinguish among various components on the basis of differences in their mass. However, even relatively mild collisional activation of ions in the ESI interface results in facile dissociation of these cluster species, yielding easily interpretable MS data (Fig. 4.9). Further optimization of experimental conditions is likely to expand the scope of direct MS measurements to include conditions that more closely mimic the *in vivo* environment.

4.1.7. Noncovalent Complexes by MALDI MS

Electrospray ionization remains the technique of choice for the detection and characterization of noncovalent complexes due to its unique ability to generate an ionic signal from protein solutions under near-native conditions. In contrast, matrix-assisted laser desorption ionization (MALDI) MS commonly relies on utilization of acidic matrices and organic solvents in sample preparation, a combination that is hardly suited to preserving noncovalent associations. In 1996, Przybylski and co-workers (61) suggested a sample preparation method that utilized a basic matrix (6-aza-2-thiothymine) without any addition of organic cosolvents, which allowed intact noncovalent protein complexes to be observed by MALDI MS. In subsequent years, several groups have reported utilization of MALDI MS to study protein–protein and protein–oligonucleotide noncovalent interactions (62–68). Nevertheless, such applications of MALDI MS are far from routine and several significant problems and questions still exist. An interested reader is referred to an informative article by Zehl and Allmaier (69), who systematically investigated several of the main problems, such as the effect of sample preparation, instrument-related effects on the stability of noncovalent complexes, as well as formation of nonspecific cluster ions.

A different approach to characterization of large noncovalent complexes by MALDI was recently introduced by Zenobi and co-workers (70), who utilized a chemical cross-linking step to stabilize complexes prior to MALDI analysis using conventional (denaturing) matrices. Since MALDI produces primarily singly charged species, the resulting mass spectra are largely devoid of complexity that is frequently seen in the ESI MS of heterogeneous samples.

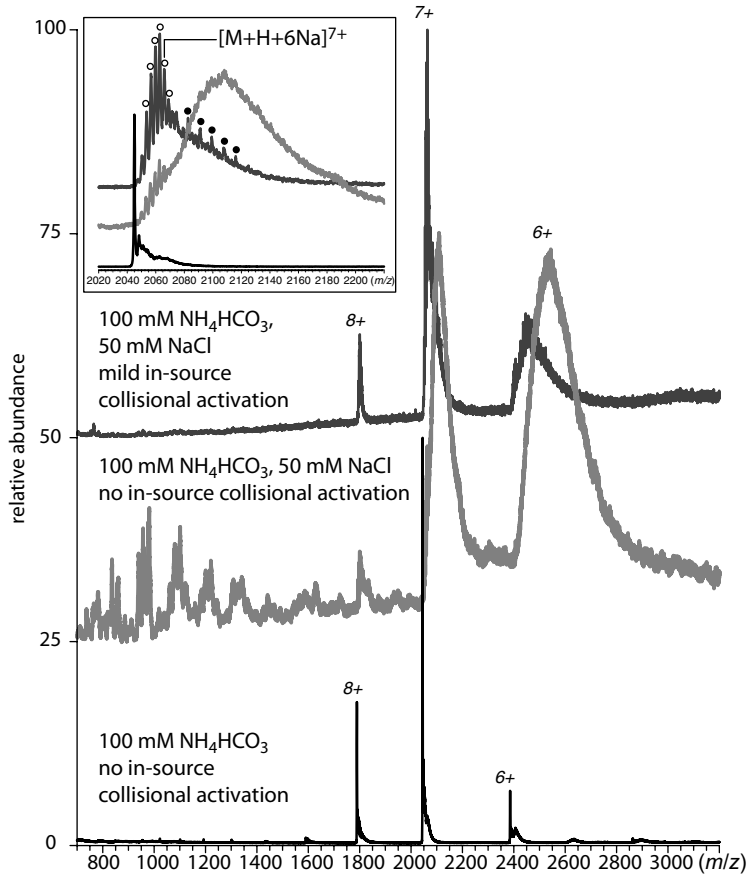


Figure 4.9. The ESI–MS of a10 μM solution of chicken egg white lysozyme acquired in 100 mM ammonium bicarbonate in the absence of nonvolatile electrolytes (black trace) and in the presence of 50 mM NaCl (gray traces). Inset shows zoomed views of lysozyme ions carrying seven charges. Peaks labeled with open circles correspond to $[M + (7 - n)H + nNa]^{7+}$ ions. Peaks marked with filled circles correspond to clusters $[M + (7 - n)H + nNa + kNaCl]^{7+}$. Adapted from (60).

The drawback of this technique is its reliance on the covalent stabilization of protein assemblies prior to MS analysis, which is very difficult to drive to completion, particularly in the case of large multisubunit complexes. Chemical cross-linking will be considered in more detail in Section 4.2. More information on using MALDI MS to study noncovalent complexes can be found in a comprehensive review (71).

4.2. CHEMICAL CROSS-LINKING FOR CHARACTERIZATION OF BIOMOLECULAR TOPOGRAPHY

4.2.1. Mono- and Bifunctional Cross-Linking Reagents

Chemical cross-linking is perhaps one of the oldest biochemical techniques used to characterize protein conformation by mapping contact topology within single proteins and their complexes. As early as the mid-1950s, Zahn and Meienhofer (72,73) used a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene, to verify the dimeric structure of insulin in solution, which was believed at the time to be a dimer formed by two stable helices spanning the entire lengths of the peptide chains. Such a view was based on hydrogen–deuterium exchange (HDX) measurements carried out earlier by Hvidt and Linderstrøm-Lang (74), who reported very high protection of the backbone amides. However, the interchain contacts detected in the cross-linking experiments did not fully support this model, leading to a suggestion that the C-terminal part of insulin is flexible in solution. As it turned out, Hvidt and Linderstrøm-Lang (75) had actually overestimated the protection of insulin due to the significant back-exchange in their initial measurements, an error that was soon discovered and corrected. This early triumph of chemical cross-linking highlighted its utility as a means of producing *proximity maps*, which can be very useful in reconstructing protein tertiary and quaternary structures at low resolution.

In the decades that followed the pioneering work of Zahn and Meienhofer (72,73), the arsenal of cross-linking reagents expanded steadily to the dozens of various bifunctional agents that are now commercially available. Cross-linking reagents are generally classified based on their chemical specificity and the length of the "spacer arm" (cross-bridge formed between the two cross-linked sites when the reaction is complete). The chemical specificity of a cross-linker determines the overall pool of reactive groups within the polypeptide that may participate in the cross-linking reaction. Only 9 of the 20 amino acid side chains are chemically active: Arg (guanidinyl), Lys (ε-amine), Asp and Glu (β- and γ-carboxylates), Cys (sulfhydryl), His (imidazole), Met (thioether), Trp (indolyl), and Tyr (phenolate) (76). Note that with very few exceptions, no reagent is absolutely group-specific. In most cases, relative chemical reactivity of a side chain is determined largely by its nucleophilicity (reactive groups of most cross-linkers interact with the side chains via nucleophilic substitution reactions). In addition to the general rules that determine nucleophilicities, the latter are affected by the state of protonation and, therefore, solution pH and microenvironment of a particular side chain. However, many reagents show a reasonable degree of selectivity (Table 4.1).

Monofunctional cross-linkers induce direct coupling of two functional groups of the protein without incorporating any extraneous material into the protein (Fig. 4.10a). Obviously, this becomes possible only if the two functional groups are in very close proximity (direct contact), in which case the cross-linker operates as a condensing agent, resulting in the cross-linked residues becoming directly interjoined. Bifunctional cross-linkers, on the other hand, contain two reagents linked through a spacer, thus allowing the coupling of functional groups whose separation does not exceed the spacer's length (Fig. 4.10b,c). Initially, most bifunctional reagents in common usage were *homobifunctional* (i.e., both cross-linking groups within the reagent targeting the same reactive groups on the protein). More recently, *heterobifunctional* cross-linkers (coupling different functional groups on the protein) have seen a dramatic surge in popularity due to a wealth of structural information that they can provide.

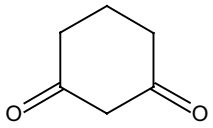
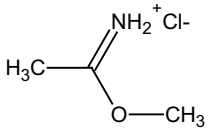
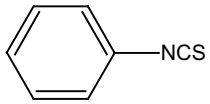
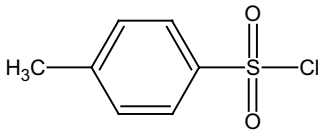
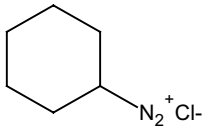
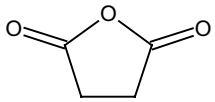
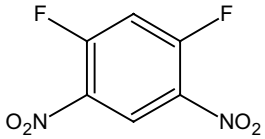
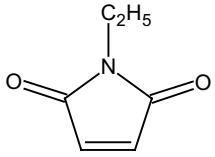
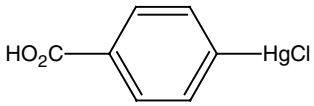
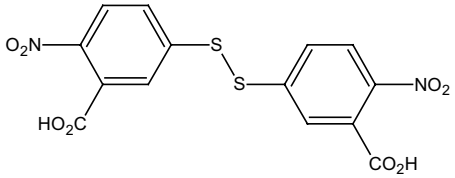
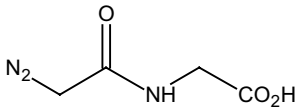
Some heterobifunctional cross-linkers may incorporate a *photosensitive* (nonspecific) *reagent* in addition to a conventional (group-specific) functionality. Such photosensitive groups react indiscriminately upon activation by irradiation. Once the specific end of such a cross-linker is anchored to an amino acid residue, the photoreactive end can be used to probe the surroundings of this amino acid (Fig. 4.10d). A variety of photo-reactive reagents can be activated at wavelengths > 300 nm, so that there is no damage to the biological macromolecule (protein or nucleic acid) due to photo-irradiation itself. Utilizing two photoreactive reagents within a cross-linker will make it totally nondiscriminatory. Trifunctional cross-linkers also have been introduced, but their utilization is mostly limited to incorporation of affinity tags in cross-linking reagents (see below). Several popular cross-linkers are listed in Table 4.2. More information can be found in several excellent reviews on the subject (78–81) and an outstanding book by Wong (76). Finally, a variety of cross-linking reagents provide a means to study higher order structure of oligonucleotides (both DNA and RNA) or oligonucleotide–protein complexes, an issue that will be discussed in more detail in Chapter 8.

4.2.2. Chemical Cross-Linkers with Fixed Arm-Length: Molecular Rulers or Tape Measures?

In structural studies, cross-linking reagents are often used as "molecular rulers," since the lengths of their spacer arms are obviously related to the distances between the cross-linked functional groups within the protein (or protein complex). If cross-linking is carried out under nondenaturing conditions, such distance constraints may provide important information regarding the tertiary and/or quaternary structural organization. A great variety of available cross-linkers provide a wide spectrum of lengths ranging from a "zero-length" reagent EDC [1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride] to several cross-linkers whose spacer arms approach 20 Å (Table 4.2). The most reliable information is provided by the zero-length reagents, as they induce formation of a direct covalent bond only between two sites that are in immediate proximity. The spacer arms of most bifunctional reagents, on the other hand, are at least somewhat flexible (the lengths presented in Table 4.2 are mean values). As a result, assigning a single number to a distance between the cross-linked sites may lead to misinterpretation of the protein structure and its dynamic character (77). Houk and co-workers (77) evaluated cross-linking "spans" using stochastic dynamics simulations for a range of homo-bifunctional reagents that target either amines or sulfhydryls. Most distributions were found to be near-Gaussian with a standard deviation ranging from 0.5 (for relatively stiff spacer arms) to 2 Å and above (for more flexible ones). However, the most striking conclusion of this work was that the distances cited for a number of popular cross-linkers, which are usually calculated as a separation between the two reactive groups in the most extended conformation of the cross-linker thereby ignoring all other possible conformations, are highly improbable.

Another important consideration that is frequently overlooked when interpreting the results of chemical cross-linking experiments is that the through-space distance between two cross-linked amino acids is often presented as the length of a Euclidean distance vector, which ignores the fact that the vector often penetrates the bulk of the protein, while the cross-linker must be physically excluded

TABLE 4.1. Commonly Used Group-Specific Reagents[a]

Class of Reagents	Example	Side Chain Specificity
Dicarbonyls		Arg
Imidoesters		Lys
Isocyanates	HNCO	Lys
Isothiocyanates		Lys
Sulfonyl halides		Lys
Diazonium salts		His, Tyr
Acid anhydrides		Lys, Tyr
Aryl halides		Cys, Lys, His, Tyr
α-Haloacetyls	ICH_2CO_2H	Cys, Lys, His, Tyr, Met
N-Maleimides		Cys, Lys
Mercurials		Cys
Disulfides		Cys
Diazoacetates		Asp, Glu, Cys

[a] Adapted from (76).

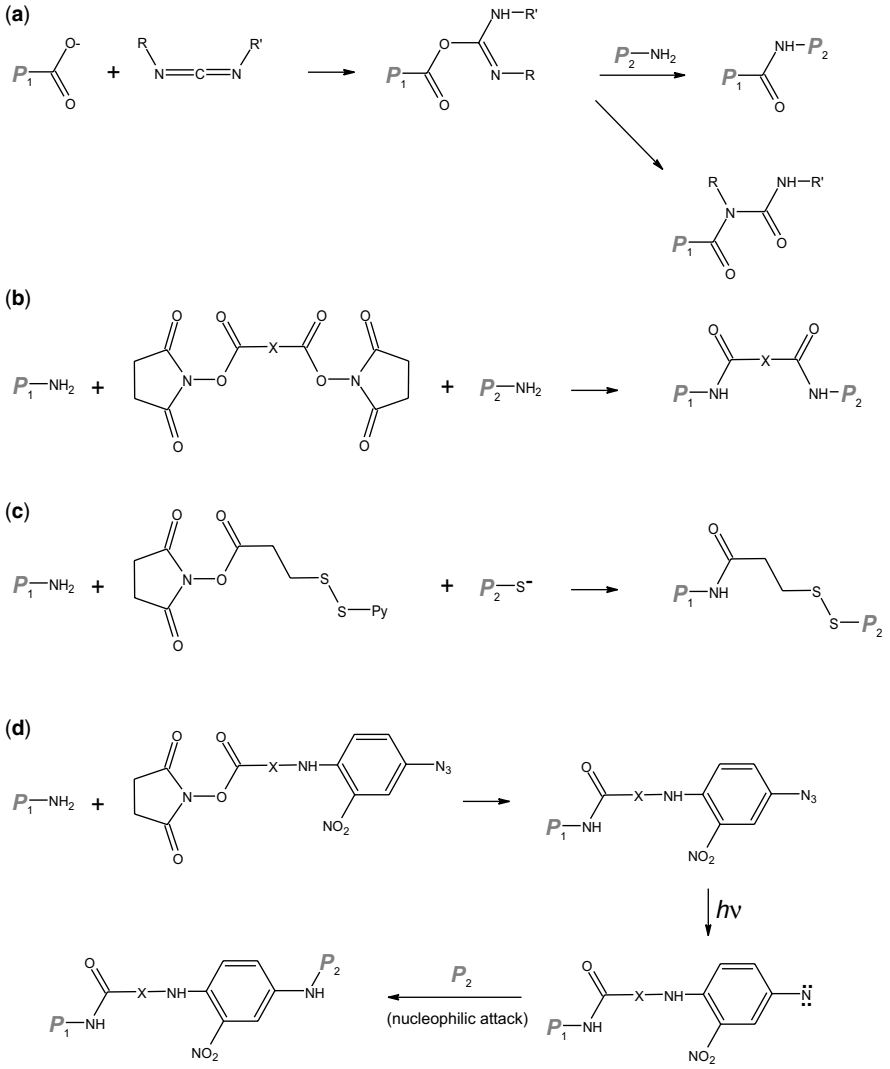


Figure 4.10. Schematic diagrams of cross-linking reactions utilizing monofunctional (*a*), homobifunctional (*b*), heterobifunctional (*c*), and photoreactive heterobifunctional (*d*) reagents.

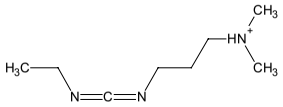
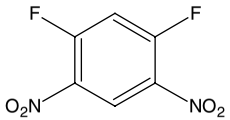
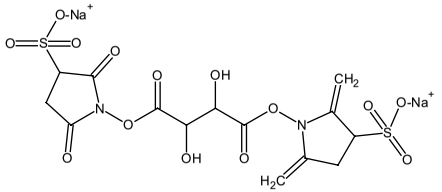
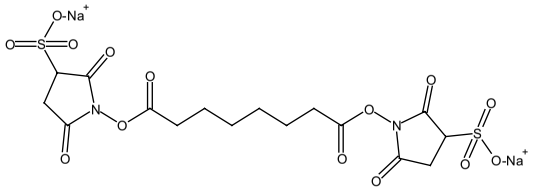
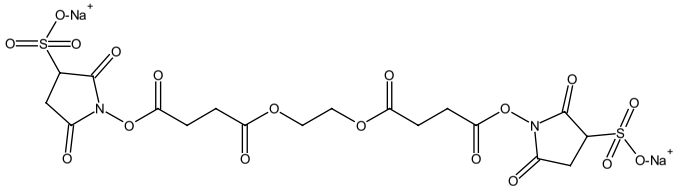
from the protein interior. This problem can be avoided using an approach where the shortest allowed distance between the cross-linked residues is calculated by taking into account protein surface geometry (82) or modeling the cross-linker onto existing protein structures explicitly (83,84). This point is illustrated in Figure 4.11, which shows that it would probably be more appropriate to regard bifunctional cross-linkers as molecular tape measures, rather than rigid molecular rulers.

4.2.3. Mass Spectrometry Analysis of Chemical Cross-Linking Reaction Products

Low-resolution cross-linking studies of protein assemblies seek only to identify the neighboring units within large protein complexes, a task that can usually be accomplished by using gel electrophoresis. Detection of the cross-linked units is often aided by using coupling reagents that incorporate radioactive labels or other reporter groups, (e.g., ultraviolet–visible (UV–vis) chromophores, as well as spin- or fluorescent labels. Mass spectrometry, on the other hand, allows quick and direct identification of the cross-linked subunits without the need to use radioactive labels or other reporter groups. The nearest-neighbor relationship for large protein complexes can be established based on the results of proteolytic digest followed by database searches to identify cross-linked proteins (85).

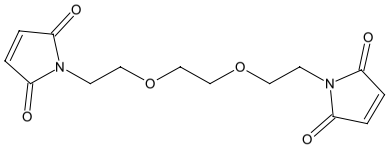
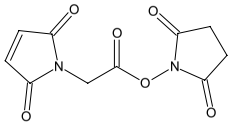
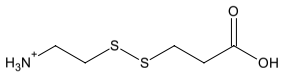
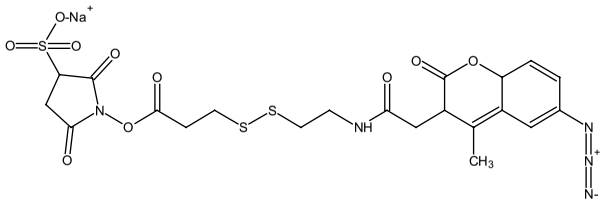
Higher resolution cross-linking studies aim at identification of the pairs of cross-linked residues within the protein or protein complex. As we discussed in the preceding section, such information may provide through-space distance constraints that are extremely valuable for defining both tertiary (intrasubunit cross-links) and quaternary (intersubunit cross-links) organization of the protein when no

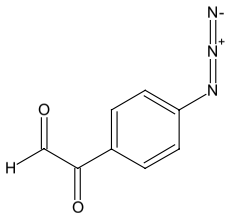
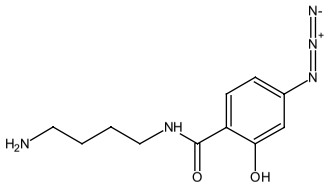
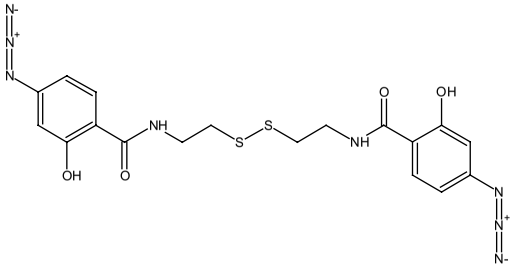
TABLE 4.2. Examples of Commonly Used Cross-Linking Reagents

Commercial Name and (Full Name)	Chemical Structure	Targeted Groups 1	Targeted Groups 2	Spacer Arm Length and Average Distance (Å)[a]	H_2O Soluble	Comments
Zero-length						
EDC 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide		$-CO_2^-$	$-NH_2$	0	Yes	
Homobifunctional						
DFDNB 1,5-Difluoro-2,4-dinitrobenzene		$-NH_2$	$-NH_2$	4.91±0.07	No	
Sulfo-DST Sulfodisuccinimidyl tartrate		$-NH_2$	$-NH_2$	6.4	Yes	
BS^3 Bis(sulfosuccinimidyl) suberate		$-NH_2$	$-NH_2$	11.4	Yes	
Sulfo-EGS Ethylene glycol bis(sulfosuccinimidylsuccinate)		$-NH_2$	$-NH_2$	16.1	Yes	

(*continued*)

TABLE 4.2 (*Continued*)

Commercial Name and (Full Name)	Chemical Structure	Targeted Groups 1	Targeted Groups 2	Spacer Arm Length and Average Distance (Å)[a]	H_2O Soluble	Comments
BM[PEO]$_3$ 1,8-Bis-maleimidotriethyleneoxide		−SH	−SH	14.7		
Heterobifunctional						
AMAS *N*-(α-Maleimidoacetoxy) succinimide ester		−SH	$-NH_2$	4.4		
AEDP 3-([2-Aminoethyl] dithio)-propionic acid		$-NH_2$	$-CO_2^-$	9.5		
SAED Sulfosuccinimidyl 2-[7-azido-4-methylcoumarin-3-acet-amido] ethyl-1,3'-dithiopropionate		$-NH_2$	Any (photo)	23.6		

APG *p*-Azidophenyl glyoxal		Guanidio *n* (Arg)	Any (photo)	9.3	No	Reacts selectively with Arg at pH 7–8
ASBA 4-(*p*-Azidosalicylamido) butylamine		$-CO_2^-$	Any (photo)	16.3		
BASED Bis(β-[4- azidosalicylamido]ethyl) disulfide		Any (photo)	Any (photo)	21.3		

[a] Reference (77).

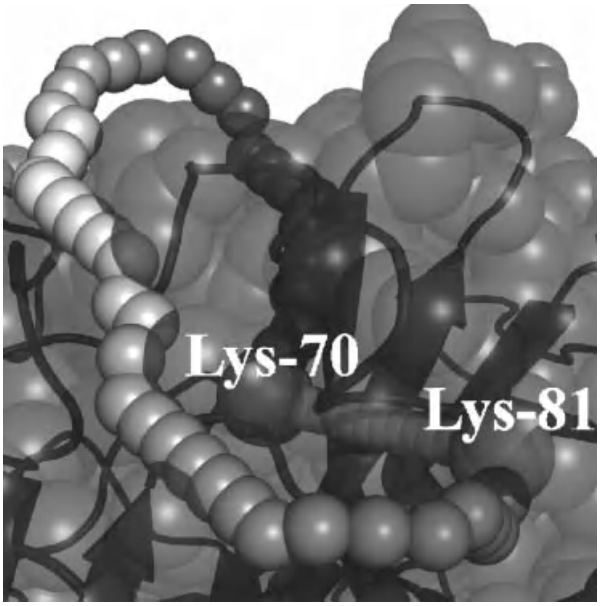


Figure 4.11. Shortest solvent accessible surface distance (SASD) path illustrated with human prothrombin. The C_β atoms of Lys-70 (orange sphere) and Lys-81 (green sphere) have a Euclidean distance of 9.1 Å (yellow vector), which value would have been in the cross-link range for DSS or BS^3 (refer to Table 4.2 for structures). However, the shortest path with an SASD of 59.2 Å reveals that the Euclidean distance vector actually penetrates the protein, leaving the only option to connect both amino acids via a long detour over the protein surface (chain of spheres colored blue to red for distances of 0 to 59 Å, respectively). [Adapted with permission from (84). Copyright © 2011 Oxford University Press.] (See color version of the figure in Color Plate section)

other structural information is available. However, confident assignment of the pairs of coupled residues within the cross-linked protein(s) is a rather challenging experimental task. A combination of proteolysis, separation methods (e.g., liquid chromatography) and MS, (and, particularly, MS/MS) provides perhaps the most elegant and efficient way of solving this problem (80,86,87). Figure 4.12 shows a workflow of a typical cross-linking experiment. Separation of proteolytic fragments prior to MS (or MS/MS) analysis usually results in significant improvements in sensitivity by eliminating possible signal suppression effects that may otherwise result in discrimination against larger (cross-linked) fragments (86). Although gel electrophoresis, high-performance liquid chromatography (HPLC), or gel permeation chromatography remain the most popular choices for sorting out products of the cross-linking reactions prior to proteolysis, utilization of capillary zone electrophoresis may result in significantly better resolution of different cross-linked species (88).

Peptide mapping alone can sometimes lead to confident identification of the cross-linked residues (89–91). For example, using trypsin as a proteolytic agent allows the quick identification of Lys residues cross-linked by homobifunctional reagents targeting this residue (89). Utilization of cleavable cross-linking reagents can also simplify data interpretation. For example, comparison of peptide maps for proteins cross-linked with reagents incorporating a thiol–thiol bond with and without prior reduction of the thiol linkers can be very helpful vis-à-vis identification of the cross-linked peptides (90). In many cases, however, unambiguous assignment of cross-linked peptides requires that MS/MS sequencing of the proteolytic fragments be carried out (92,93). A nomenclature for MS/MS analysis of cross-linked peptides introduced by Gibson and co-workers (94) categorizes both different types of cross-linked peptides (products of chemical cross-linking followed by proteolysis) and gas-phase fragmentation patterns. Peptides are classified according to the outcome of the cross-linking reaction, assigning type 0 for the dead-end products (see below), type 1 for internal cross-links, and type 2 for inter-peptide cross-links (Fig. 4.13).

The MS/MS analysis of the cross-linked species frequently provides advantages in speed, sensitivity, and capability of handling large protein assemblies. However, heterogeneity of the peptide mixture and relatively low yields of the cross-linked products frequently make the detection and confident assignment of the cross-linked peptides challenging (95). Incorporation of affinity tags into cross-linking reagents (e.g., by utilizing trifunctional reagents, where one of the reactive sites is conjugated to biotin) alleviates the problem related to low yields of cross-linking reactions to a significant extent (80). However, many popular cross-linking reagents (especially the ones targeting Lys side chains) are prone to degradation via hydrolysis, leading to formation of dead-end products (covalently attached to ε-amine of one Lys residue, but unable to bind a second amino group). Predominance of these species among the cross-linking reaction products cannot be dealt with using affinity tags alone, but can be addressed by incorporating isotope tags into the cross-linking reagents (96). Both the isotope and the affinity tags can be combined in a single cross-linking reagent (95). In many cases, isotope tagging not only simplifies the assignment of the cross-linked residues in the MS/MS spectra (96), but also improves the detection limits and even allows parallel use of two cross-linkers of similar reactivity, but different spacer length (97). Another approach to developing cross-linking reagents that is particularly well suited to MS and MS/MS makes use of signature ions, stable fragment ion markers that are readily produced upon collisional activation of cross-linked peptide ions (98).

Interpretation of the MS and MS/MS data on cross-linked peptides is greatly simplified as the resolution and mass accuracy of MS measurements increase (99). Identification of cross-linked sites can also be facilitated by using digestions by multiple protease in combination with MS and

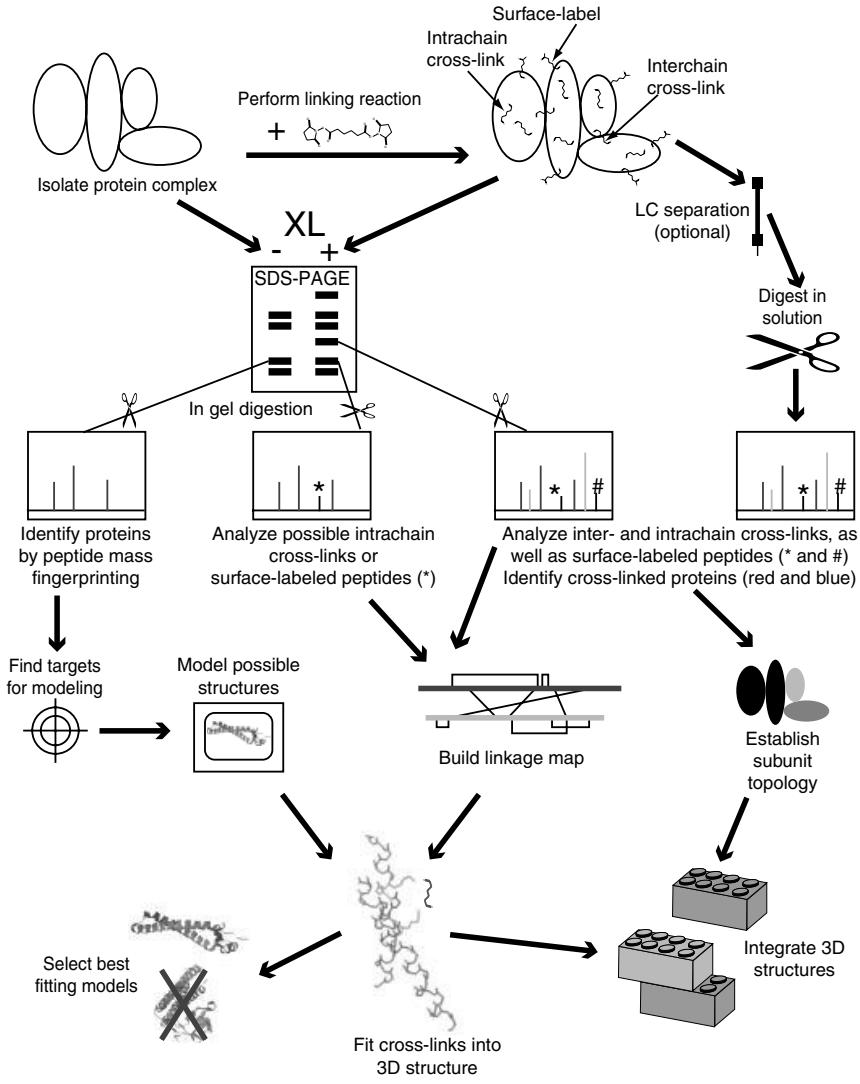


Figure 4.12. A schematic diagram of workflow of cross-linking a multiprotein complex and integrating the levels of information into a 3D model of the structure. [Re-printed from (86). Copyright © 2003 with permission from Elsevier Ltd.]

MS/MS analysis of the proteolytic fragments (100). Finally, rapid proliferation of MS instruments capable of carrying out electron-based fragmentation (ETD or ECD) has also benefited the field. In many cases, ECD and ETD provide more complete sequence coverage of cross-linked peptides, allowing the modified residues to be localized with higher precision. Additional benefits can be provided by utilization of specialty cross-linking reagents, such as 1,3-diformyl-5-ethynylbenzene, (DEB), which converts ε-amino groups of Lys residues to secondary amines (101). Unlike conventional reagents, DEB does not eliminate the ability of modified Lys residues to carry protons, thereby increasing the charge state of the cross-linked peptide and enhancing its ETD efficiency, and also allowing the selection of cross-linked peptides to be made based on the charge state. In addition to facile backbone cleavage, ETD of peptide–DEB–peptide conjugates produces diagnostic ions that reveal the identity of the unmodified peptides (101).

As the amount of information deduced from cross-linking experiments increases, so does the complexity of data interpretation, and the tools of bioinformatics become absolutely essential for interpretation of the results of cross-linking experiments (102). The task of assigning the cross-linked peptides and localizing the modification sites can be greatly assisted by a variety of automated algorithms that use MS or MS/MS data as input (86,102). The database mining approach to identification of cross-linked peptides mentioned earlier in this section (85) can be used even in a situation when the protein complex composition is not known *a priori* (103). More sophisticated approaches, such as Xlink-Identifier (104), allow the cross-linking sites to be localized with high precision by identifying inter- and intra-peptide cross-links

in addition to dead-end products and underivatized peptides. Another comprehensive cross-linking data analysis platform is MS-Bridge (105), which is part of the Protein Prospector MS data analysis suite. While these platforms were developed to support label-free analyses, several other algorithms have been developed to take advantage of isotopically tagged cross-linkers (106–109). A comprehensive list of data analysis programs developed for interpretation of the results of cross-linking experiments can be found in a recent review article (87).

4.2.4. Intrinsic Cross-Linkers: Methods to Determine Disulfide Connectivity Patterns in Proteins

While the thiol groups of cysteine residues in cellular redox-active proteins undergo frequent oxidation and reduction based on their functional state (110), disulfide patterns in most other proteins are highly conserved. These natural cross-links serve the purpose of reinforcing the tertiary and, in some cases, quaternary structure of a large number of proteins, and any change in the thiol–thiol connectivity pattern (commonly referred to as disulfide scrambling) almost always leads to conformational changes. Therefore, disulfide patterns can serve as reporters of integrity of the higher order structure of proteins. The intimate involvement of disulfide bonds in maintaining conformation has also led to suggestions that connectivity patterns of cysteine residues can be used to identify protein homologues that are otherwise undetectable (111), and can even aid in prediction of the higher order structure of disulfide-rich polypeptides in the absence of structural homologues (112). Since these predictive algorithms are still in their infancy, this section will focus our attention on using disulfide mapping to monitor protein conformational integrity.

Although the protein mass remains unchanged as a result of disulfide scrambling, a combination of liquid chromatography (LC) and MS may allow the alteration of cysteine linkage patterns to be detected by observing multiple LC peaks, each corresponding to the same protein mass (113). Rearrangement of disulfide bonds in a protein may also result in significant changes of its conformation in the gas phase, which in some cases can be detected using ion mobility MS (114). Localization of scrambled disulfides usually requires a significant effort, particularly in cysteine-rich proteins, and peptide mapping typically must be complemented by tandem MS analysis. Since CAD in the positive ion mode leaves thiol–thiol linkages in peptides intact, other ion activation methods are frequently employed. Electron capture dissociation (ECD) and negative ion CAD (115–117) are particularly useful, since they both have the ability to cleave disulfide bonds. The latter technique is particularly appealing, since it results in very inefficient backbone cleavage, thereby allowing the disulfide-linked peptide dimers to be identified in complex

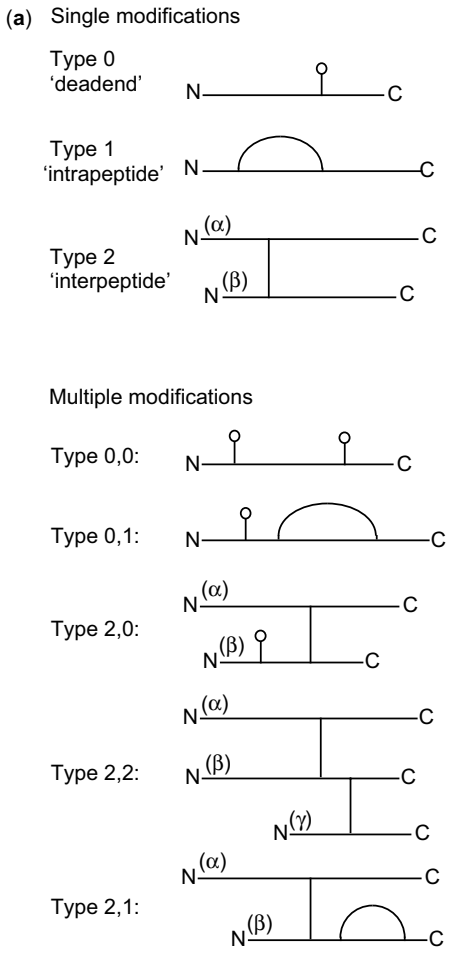


Figure 4.13. Classification of cross-linked peptides and nomenclature of fragment ions derived from the cross-linked peptides. (*a*) Classification of cross-linked peptides into *type 0*, *type 1*, and *type 2* outcomes and an extension of this nomenclature to combinations of two outcomes for cases of multiple cross-linking and/or modification. Chain length or mass ($\alpha > \beta$) and sequence position (N- to C-terminus) determine the order of the two numbers that designate the type of a cross-link). (*b*) examples of fragmentation of the *type 1* peptide involving a *y*, *b*-cleavage (e.g., y_6b_6) or alternatively, the loss of an amino acid from the cyclic portion of the cross-linked peptide (e.g., $-AA_4$ or y_3b_3 cleavage reaction). (*c*) examples of fragmentation of the type *2* peptides. [Adapted with permission from (94). Copyright © 2003 American Society for Mass Spectrometry.]

mixtures (118). Determination of disulfide patterns in cellular redox-active proteins is more difficult, as it requires that special care be taken not to upset the delicate balance among various oxidation states of cysteine, which is reflective of a particular functional state.

In addition to disulfide scrambling, there is another, less frequently occurring covalent modification that involves cystine–cystine bonds. The so-called trisulfides are formed by insertion of a sulfur atom into a disulfide bond (119). Trisulfide formation appears to be a common modification

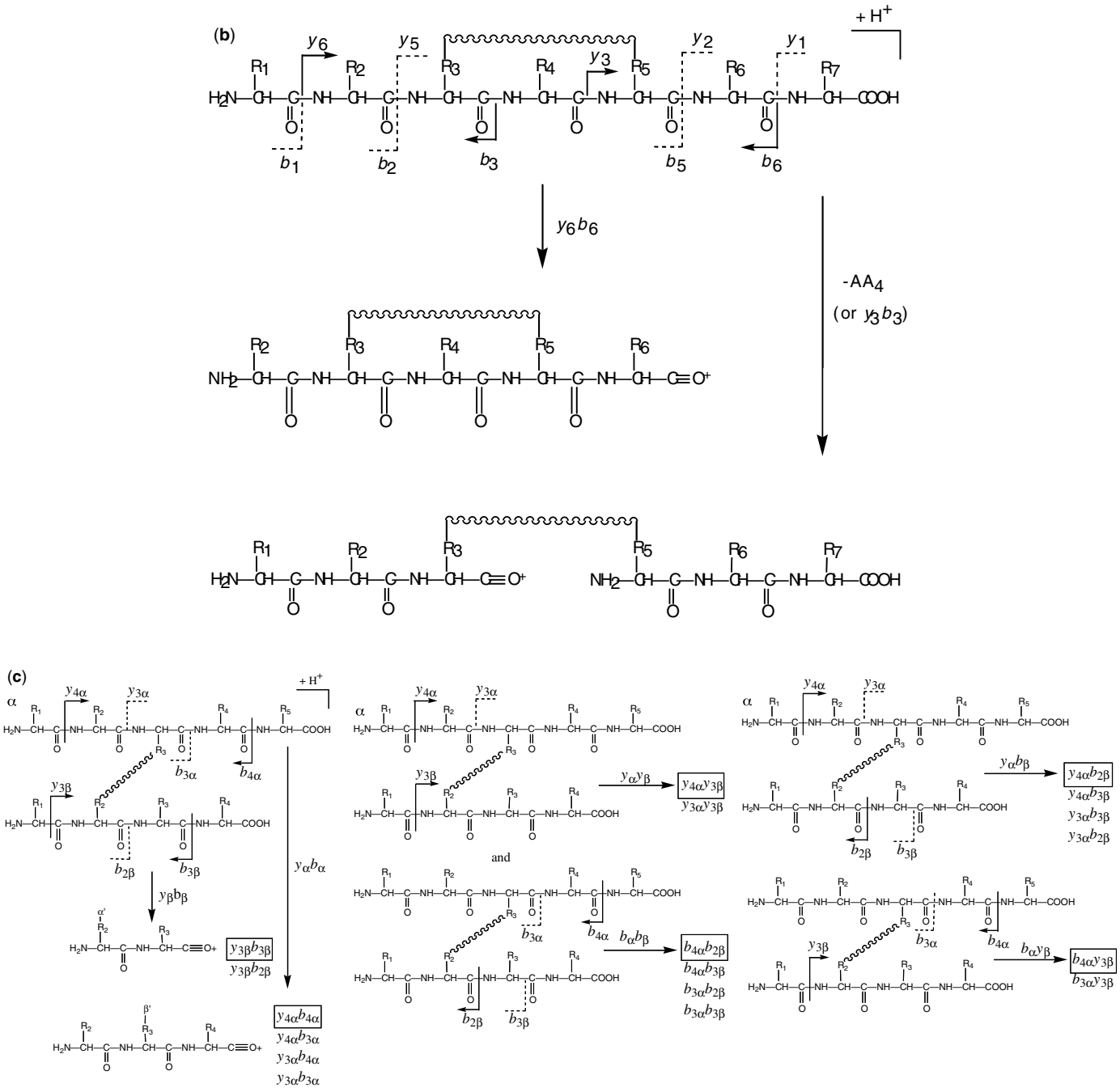


Figure 4.13. (*Continued*)

in many natural and recombinant antibodies, where it occurs only in interchain linkages (120), although several instances of conversion of internal disulfide bonds to trisulfides have also been documented for monomeric proteins (119). Detection of these modifications is relatively straightforward, as trisulfide formation results in a noticeable mass increase not only at the peptide level (120), but also at the protein level (121). More information on various MS based methods to map disulfide bonds can be found in a recent review on the subject (122).

4.2.5. Other Intrinsic Cross-Linkers: Oxidative Cross-Linking of Tyrosine Side Chains

Finally, a very interesting approach to carrying out cross-linking experiments utilizes a minimally invasive procedure that makes use of intrinsic reactivity of side chains of amino acids other than cysteine. For example, a dityrosine cross-link is frequently encountered in nature (particularly in structural proteins e.g., elastin and silk), where its formation

is catalyzed by certain metalloenzymes. However, dityrosine formation can also occur *in vitro*, where it is catalyzed by metal–peptide complexes (123). This feature was exploited by Craik and co-workers (124), who suggested using the short nickel-binding peptide GGH as an *intrinsic reagent* that can be expressed as part of a protein (124,125). In the presence of Ni^{2+} salts and oxidants (e.g., oxone and monoperoxyphthalic acid) this peptide supported cross-linking of the two subunits of a homodimeric enzyme with no observable protein degradation, and could also cross-link the enzyme to its physiological targets. The yields of the cross-linking reaction, which was presumed to proceed via formation of bityrosyl cross-links appeared to be sensitive to the inter-tyrosine distance in the protein structure (125).

However, later studies revealed that this cross-linking reaction does not necessarily proceed via forming a covalent linkage between two tyrosine residues, and may in fact be relatively nonspecific (100). Subsequently, most recent studies exploiting tyrosine reactivity have reverted to a strategy where the Ni-binding peptide is not a part of the protein sequence, but is introduced into the system alongside other reagents. This latter approach was used successfully (in parallel with photocross-linking) to probe quaternary structure of cowpea mosaic virus capsid, which is composed of 60 identical copies of a two-subunit protein (126). Nickel-catalyzed intersubunit cross-linking in this study was found to occur exclusively at adjacent tyrosine residues, as predicted from the X-ray crystal structure of the capsid. (126).

4.3. MAPPING SOLVENT-ACCESSIBLE AREAS WITH CHEMICAL LABELING AND FOOTPRINTING METHODS

4.3.1. Selective Chemical Labeling

Selective chemical modification (127) is another technique that has enjoyed increasing popularity in a variety of biophysical studies. Although initially developed primarily as a tool to modulate enzymatic activity (128–130), its utility for studying protein–protein interactions by mapping contact interfaces was recognized early on (131). However, it was not until Fenselau and co-workers (132) demonstrated the power of MS as a means of localizing both shielded and modified residues that this technique became an efficient probe of higher order macromolecular structure. The extent of chemical modification of a certain amino acid residue is used to determine its solvent accessibility. When collected in a residue-specific fashion, such information can be used to map solvent-exposed areas of the protein (Fig. 4.14). The most popular amino acid-specific modifiers are shown in Table 4.3, and a larger list of reagents used in selective chemical labeling experiments can be found in an excellent review by Mendoza and Vachet (133). Any chemical modification of an amino acid side chain alters the protein mass, hence the appeal of MS as a readout tool for the outcome of such experiments. Interpretation of the MS and MS/MS data on chemically modified proteins is usually relatively straightforward (as compared to the analysis of cross-linked proteins) and greatly benefits from a vast arsenal of experimental tools developed to analyze post-translational modifications of proteins.

In a typical experiment, modified proteins are processed with a suitable proteolytic enzyme, followed by mass mapping of the fragment peptides. The position(s) of the modified residue(s) within each proteolytic fragment can be reliably established using MS/MS, as the presence of chemical modification manifests itself as a break or a shift in the ladder of the expected fragment ions. Intersubunit binding topology is usually determined by comparing modification patterns of the protein obtained in the presence and in the absence of its binding partner (134), although the two experiments can be combined if the labeling agent contains a stable isotope tag (135). An added benefit of using isotope tags is the easy recognition and quantitation of label-containing peptides and their fragments in MS and MS/MS spectra. Finally, incorporation of a chromophore into a

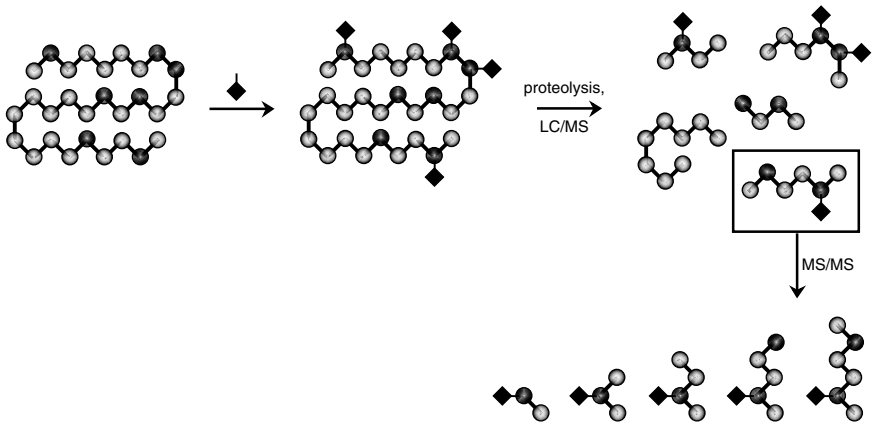
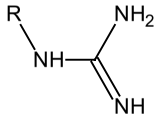
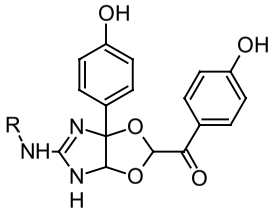
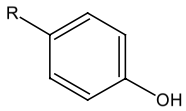
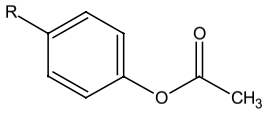
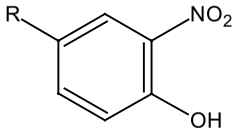
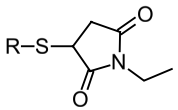


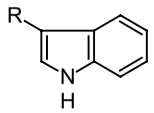
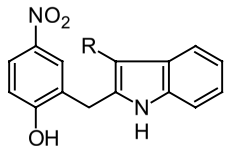
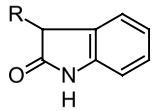
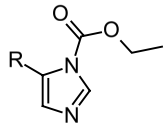
Figure 4.14. Mapping solvent accessible areas of protein using selective chemical modification.

TABLE 4.3. Examples of Commonly Used Reagents for Selective Chemical Modification of Amino Acid Side Chains in Proteins

Amino Acid	Reactive Group	Reagent	Side Chain Derivative	Comments
Lys	**R–NH$_2$**	Acetic anhydride	**R–NH**–CO–CH$_3$	Large excess of the reagent is added stepwise; optimal solution conditions pH 7–9.5, 0°C
		Cyanate	**R–NH**–CO–NH$_2$	Optimal reaction conditions pH 8, 37°C (12–24 hs); side reactions involve reversible modification of –SH and –C$_6$H$_5$OH groups
Arg		HPG (*p*-Hydroxyphenyl glyoxal)		Reacts very specifically under mild conditions (pH 7–9, 25°C)
Tyr		Acetyl imidazole		Optimal pH 7.5–8.0; side reactions involve acetylation of –NH$_2$ and –SH groups (acetylation of serine residues is also possible)
		TNM (tetranitromethane)		Optimal pH 8.0; side reactions involve modification of Met and Trp; cross-linking and polymerization are also possible (see Section 4.2.1)
Asp, Glu	**R–CO$_2$H**	Diazoacetamide	**R–CO–O**–CH$_2$CO–NH$_2$	Requires Cu^{2+} as a catalyst; optimal conditions pH 5, 15°C for 1–2 h
		EDC + glycineamide	**R–CO**–NH–CH$_2$CO–NH$_2$	Optimal pH 4.5–6, reaction time 0.5–2 h
Cys	**R–SH**	Iodoacetic acid	**R–S**–CH$_2$–CO$_2$H	Cross-reactivity with Cys and (to a lesser extent) Tyr Modification of Met, Lys, and His may also occur
		N-Ethylmaleimide (NEM)		More specific, but still cross-reacts with Met May undergo partial ring hydrolysis

(*continued*)

TABLE 4.3 (*Continued*)

Amino Acid	Reactive Group	Reagent	Side Chain Derivative	Comments
Trp		2-Hydroxy-5-nitrobenzyl bromide (Koshland reagent)		Highest specificity for Trp at pH 4–5 (cross-reacts with Cys at pH < 4 and Tyr at alkaline pH) Very sensitive to hydrolysis; photosensitive (reactions must be carried out in the dark)
		N-Bromosuccinimide (NBS)		Highest specificity for Trp at pH 5 Labeling is reversible
His		Diethyl polycarbonate (DEPC)		Optimal pH 5.5–7.5 High concentration of DEPC causes second modification of His side chain Cross-reacts with Lys, Tyr, Cys, Arg, Ser, and Thr

reagent provides a parallel route of detection of the chemically modified species, which can be used to monitor the progress of selective chemical modification, as well as to quickly identify label-carrying fragments following proteolytic digestion of the modified protein (136).

Although generally regarded as a low-resolution technique, selective chemical modification can sometimes provide structural information at surprisingly high resolution. One particularly intriguing example was reported by Hager-Braun and Tomer, (134) who used a combination of selective chemical modification, proteolysis, LC, and MS/MS to study interaction between the full-length human immunodeficiency virus (HIV) gp120 protein and a soluble segment of human CD4 (134), a receptor playing an intimate role in granting HIV access into T cells. Of the two adjacent Arg residues located at the CD4–gp120 interface (Arg^{58} and Arg^{59}), only the latter becomes shielded from the solvent in the presence of gp120, while both residues react with hydroxyphenylglyoxal (HPG) in the absence of the viral protein (Fig. 4.15). Therefore, only Arg^{59} (but not its nearest neighbor) was established as a part of the CD4–gp120 interaction site.

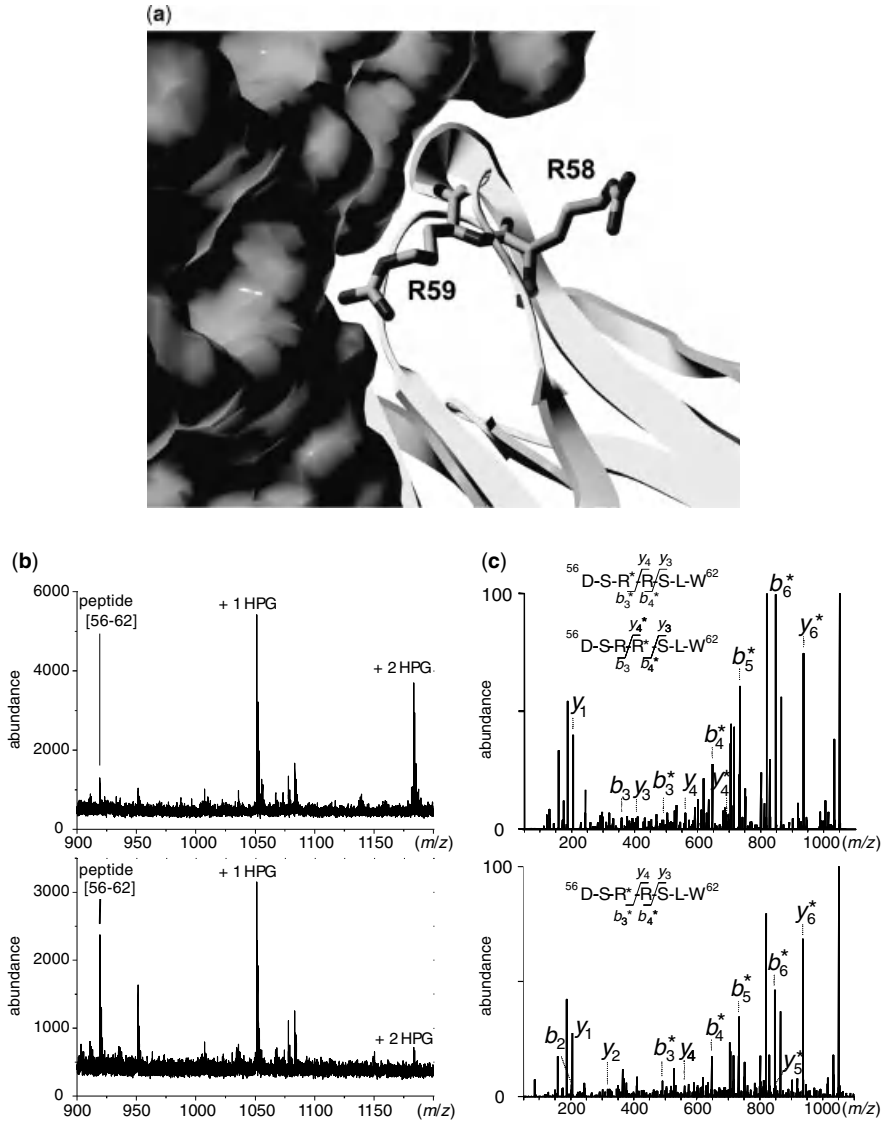


Figure 4.15. (*a*) interaction site of CD4 (ribbon) with a truncated HIV envelope protein gp120 (surface) derived from the crystal structure. The Arg^{59} of CD4 interacts with gp120, while Arg^{58} does not (both arginine residues depicted as sticks). (*b*) The MALDI mass spectra of proteolytic peptides (AspN) derived from CD4 modified with HPG in the absence (top) and in the presence of gp120 (bottom). (*c*) Localization of the modified arginine residue within the proteolytic peptide (56–62) using MS/MS. [Reprinted from (134).]

An even more intriguing suggestion regarding further improvements in spatial resolution in chemical modification experiments was made by Pucci and co-workers (137). They argued that modification of tyrosine residues with reagents targeting different functional groups (i.e., *O*–acetylation of the hydroxyl group with *N*-acetyl imidazole (138) and modification of the aromatic ring with tetranitromethane (TNM) (139,140) may provide information on the microenvironments of these functional groups, thereby extending the spatial resolution beyond the level of a single amino acid. When the two reagents were used to probe solvent accessibility of six Tyr residues of Minibody (a small *de novo* designed β-protein), only three were modified by *N*-acetyl imidazole, while five were reactive toward TNM (137). These observations led to a conclusion that the two extra Tyr residues (the ones that can be modified only by TNM) are only partially accessible to the solvent, with their phenolic hydroxyl groups being involved in hydrogen bonds (137). However, any conclusion based solely on the outcome of chemical modification of the protein with *N*-acetyl imidazole should be met with certain skepticism, as Tyr acetylation experiments are known to be rather uninformative as far as position of the phenolic hydroxyl groups (see below).

Although the interpretation of the results of chemical modification experiments is often based on the assumption that reactivity of side chains is correlated to their location within the protein, this is not universally true. Soon after the first high-resolution structures of several proteins became available, it became clear that assignment of buried and exposed groups within a native protein based solely upon chemical modification data may be incorrect in any given instance (141). Acetylation of Tyr with *N*-acetyl imidazole mentioned in the previous paragraph seems to be a particularly tricky case, since the solvent-exposed residues often show little or no reactivity [a review of such cases is presented in (141)]. Obviously, functional groups can remain sterically hindered even at the surface of the protein. For example, Liu and Reilly (142) recently demonstrated that Lys residues located on the protein surface remain unlabeled if they are proximal to acidic residues, most likely due to formation of salt bridges that prevent their chemical modification.

Other factors may also play an important (and sometimes even decisive) role in modulating reactivity of amino acid side chains. One particularly interesting example is modification of Arg side chains with 1,2-cyclohexanedione, CHD (143,144). Differential reactivity of Arg residues toward CHD is often interpreted as resulting from varying solvent accessibility (145). However, Przybylski and coworkers (146) used a model protein lysozyme to demonstrate that the reactivity of Arg residues toward CHD did not correlate with their solvent accessibility. In fact, these experiments showed an inverse correlation, with the two most reactive Arg residues having the lowest solvent accessibility factors. Interestingly, acetylation of Lys residues carried out on the same protein yielded a more intuitive reactivity order in agreement with the solvent accessibility factors calculated based on the known crystal structure of the protein. This intriguing observation has been explained in terms of intramolecular catalysis (by a nearby proton acceptor group) as being a prerequisite for a successful modification of the guanidine group. Such apparent importance of chemical microenvironment on the efficiency of modification reactions highlights difficulties associated with interpretation of the experimental results.

In many other cases, reasons why a certain functional group fails to react with its specific reagents are more obscure. Furthermore, a surprisingly high reactivity is sometimes displayed by buried side chains, which are sequestered from the solvent according to crystal structures. Such anomaly can be caused by a variety of factors, including significant alteration of the side chain pK_a in the protein interior, catalytic effect of neighboring groups (as discussed above), efficient trapping (binding) of the reagent inside the protein in close proximity to the modification site, or highly dynamic character of the protein segment in question. This latter point is best illustrated with our own recent work, where modification of interferon-β1a with *N*-ethylmaleimide (NEM) results in efficient labeling of a Cys residue that resides in the protein interior (147). Subsequent studies of protein dynamics with hydrogen–deuterium exchange demonstrated that the segment containing the buried Cys residue is highly dynamic, suggesting that its modification can occur during transient exposure of the thiol group to solvent (147).

Finally, an important consideration that must be kept in mind when designing or interpreting the results of selective chemical-labeling experiments relates to the fact that structural information derived from such measurements is reliable only if the protein maintains its conformation during the experiment (133). Most chemical modifications result in changing the charge of the labeled amino acid residue, and significant alteration of the protein surface charge distribution may obviously result in conformational change. Even the sheer size of many groups used as covalent labels may interfere with the protein's ability to maintain its conformation by creating steric constraints: in the example mentioned above, modification of interferon-β1a with a single NEM group resulted in significant destabilization of tertiary structure and ~90% loss of biological activity (147). Despite the extreme seriousness of this concern, less than one-half of all studies utilizing selective chemical labeling that were conducted in the past decade employed any means of ensuring the integrity of protein higher order structure during the experiments (133).

Artifacts associated with the influence of chemical modifications on the protein conformation can be avoided by limiting the number of modifications to one per protein

molecule. In this way, reactivity of any amino acid side chain is determined only by the unperturbed protein structure (133). However, even though this approach can be implemented in the case of oxidative labeling (discussed in Section 4.3.2), often it is unpractical when dealing with selective chemical labeling, as this would result in very low yields of chemical modification products. Awareness of the problems that may arise in the course of selective chemical modification studies is essential. However, the usefulness of such studies in assessing protein higher order structure is unquestionable.

4.3.2. Nonspecific Chemical Labeling

While chemical modification studies employing selective (amino acid specific) reagents continue to be used in structural studies, approaches utilizing nonspecific labeling have seen a truly dramatic surge of popularity in recent years. A significant advantage offered by this approach is universality, that is, the ability to determine the solvent exposure of several types of amino acids simultaneously in a single experiment. So far, hydroxyl radical $HO^{\bullet}$ is the most popular nonspecific modifier, due to its ability to induce side-chain oxidation for a variety of amino acids, and the relative ease of its generation in solution. The $HO^{\bullet}$ radical can be generated chemically (148–150), electrochemically (151), photochemically (152) or radiolytically (153–155).

The hydroxyl radical is relatively nondiscriminatory, and can modify virtually all types of amino acid side chains, including aliphatic (via direct insertion of a hydroxyl group), aromatic (e.g., via direct insertion of a hydroxyl group into the aromatic ring of Tyr and Trp), basic (e.g., by replacing the guanidine group with an aldehyde group in arginine and converting lysine to hydroxylysine), and acidic via decarboxylation (156). However, the most susceptible to $HO^{\bullet}$ attack are side chains containing sulfur atoms (Cys and Met), including disulfide-bonded Cys residues. The least susceptible to $HO^{\bullet}$ attack are Gly, Asn, Asp, and Ala, whose reactivity is three orders of magnitude lower than that of Cys. Although side chains are the prime targets of attack by hydroxyl radicals, main chain cleavage may also occur (156). The great variety of $HO^{\bullet}$ induced oxidation products and the large number of potential targets places a premium on the ability to detect and identify the modification sites. Usually, proteolytic degradation of the modified protein followed by LC–MS and MS/MS analyses is needed in order to achieve reliable identification of oxidatively labeled amino acid side chains (156–158).

Although oxidative labeling is nonspecific, the significant range of $HO^{\bullet}$ radical reactivity to different side chains inevitably limits the ability to observe oxidation of certain residues (e.g., Gly, Asn, Asp, and Ala mentioned above). Additional complication arises due to the prevalence of multiple oxidation products for several amino acids, which further reduces the sensitivity of the analysis. Kiselar and Chance (159) note that only about one-half of naturally occurring amino acids are used routinely as probes in oxidative labeling experiments, which obviously has a negative impact on the spatial resolution afforded by such measurements. However, this problem can be circumvented to some degree by using the knowledge of modification chemistry to predict the LC elution time of modified peptides, and advanced MS/MS methods, such as selected reaction monitoring (159). As is the case with the analysis of the results of chemical cross-linking experiments, extracting useful information from the covalent labeling experimental data greatly benefits from automation. Chance and co-workers (160) recently introduced ProtMapMS, a suite of algorithms developed specifically for the comprehensive automatic analysis of MS data obtained for protein structure studies using covalent labeling. The functions of ProtMapMS include data format conversion, MS data interpretation, detection and verification of all peptide species, confirmation of the modified peptide products, and quantification of the extent of peptide modification.

In Section 4.3.1, we mentioned that selective chemical labeling can result in artifacts, since progressive decoration of the protein with chemical labels will eventually lead to conformational changes. The same is true for oxidative labeling; in fact, $HO^{\bullet}$ induced modifications have been shown in many cases to trigger changes in protein higher order structure, as reviewed in (157). Therefore, it is desirable to keep the extent of protein modification low to minimize conformational perturbation (157), although this inevitably has a negative impact on the sensitivity of the measurements. A very elegant solution to this problem was proposed by Hambly and Gross (161), who noted that the extent of artifacts introduced by chemical labeling depends not only on the extent of protein oxidation, but also on the time frame of oxidation. Should this reaction time window be significantly narrow compared to the submillisecond time scale of conformational changes, the labeling pattern would reflect only the native structure of the protein, even if the number of modified sites on each protein is significant. These considerations form the basis of a highly successful technique called fast photochemical oxidation of proteins (FPOP), where solvent-exposed amino-acid residues are oxidized by $HO^{\bullet}$ radicals produced by the photolysis of H_2O_2. FPOP is designed to limit protein exposure to radicals to $< 1\,\mu s$ by employing a pulsed laser for initiation to produce the radicals and a radical-scavenger to limit their lifetimes (152).

While most nonselective chemical labeling studies utilize the $HO^{\bullet}$ radical, other probes can be used as well. For example, it was shown recently that the FPOP scheme can accommodate $SO_4^{\bullet-}$(generated by 248 nm homolysis of persulfate) as a radical reactant. Another popular photoactivated nonspecific labeling reagent that is frequently used

in structural studies is diazirine, CH_2N_2, which generates highly reactive methylene carbene upon photolysis (162–165), as well as other diazirine-based modifiers (166).

4.4. HYDROGEN EXCHANGE

4.4.1. Hydrogen Exchange in Peptides and Proteins: General Considerations

Hydrogen–deuterium exchange (HDX) is a general experimental technique that detects the presence or absence of hydrogen bonding within a protein in solution (167). The analytical value of HDX as a tool for probing macromolecular structure was recognized almost immediately after the discovery of deuterium (168) and subsequent development of the methods of production of heavy water (169). Initial studies of the exchange reactions between organic molecules and ^2H_2O carried out by Bonhoeffer and Klar (170) indicated that that the exchange rate is very high for heteroatoms (e.g., $-$OH groups), while the hydrogen atoms attached to carbon atoms (e.g., $-CH_3$ groups) do not undergo exchange. Hvidt and Linderstrøm-Lang (74,75) later used HDX to measure solvent accessibility of labile hydrogen atoms as a probe of polypeptide structure. While this work was the first attempt to use HDX to probe protein structure in solution, the early studies of hydrogen-exchange reactions between water and biopolymers preceded this work by more than 15 years (177). Burley et al. (172) suggested that the extent of ^2H incorporation into a protein molecule can be measured by monitoring its mass increase. However, it was not until several decades later that MS was used to determine the number of labile hydrogen atoms within a polypeptide (173). The advent of ESI and MALDI MS dramatically expanded the range of biopolymers for which the extent of ^2H incorporation could be measured directly under a variety of conditions (174). As a result, HDX MS has now become a powerful experimental tool for probing protein higher order structure.

HDX targets all labile hydrogen atoms (i.e., those attached to nitrogen atoms at the backbone amides and heteroatoms at polar/charged side chains). Therefore, HDX may be viewed as a nonselective chemical modification technique, whose major difference from those discussed in the preceding sections is the reversible character of modification. Protein HDX involves two different types of reactions: (*1*) reversible protein unfolding that disrupts the hydrogen-bonding network and/or exposes buried segments to solvent, and (*2*) isotope exchange at individual unprotected amides or side chain heteroatoms. As a result, the overall HDX kinetics may be very complicated and will be discussed in detail in Chapter 5. An important feature of HDX that is particularly relevant within the context of this chapter is that the exchange rate of labile protons that are inaccessible to solvent is usually very slow. This feature has been used widely in recent years to probe protein conformation and identify binding interfaces in protein–protein complexes [reviewed in (175–179)]. The basic premise of the analysis is that HDX rate reduction for labile hydrogen atoms in a given protein segment within a complex indicates that this segment is located at the protein-protein binding interface. Analogously, HDX can be used to identify changes in protein tertiary structure induced by small ligand binding (180).

4.4.2. Probing Exchange Patterns with HDX MS at the Local Level

Replacement of each proton with a deuteron (or *vice versa*) results in a protein mass change of ~ 1 Da, which makes MS a very sensitive and reliable detector of the progress of protein HDX reactions. Mass measurements of proteins undergoing HDX are usually carried out following rapid acidification of the protein solution to pH 2.5–3 and lowering the temperature to 0–4°C, which results in significant deceleration of the chemical (intrinsic) exchange rates of backbone amide hydrogen atoms (see Chapter 5 for more detail). These conditions, known as "HDX quenching" or "slow exchange conditions", also result in unfolding of most proteins. Since the intrinsic exchange rates of labile side-chain hydrogen atoms are not decelerated as significantly as those for backbone amides (181), all information on the side chain protection is generally lost during this step, leaving a single HDX reporter for each amino acid residue (with the exception of proline, the only natural amino acid lacking a backbone amide hydrogen atom). Another fortunate consequence of quench-induced protein denaturation is dissociation of all noncovalently bound ligands (ranging from metal cations and small organic molecules to other biopolymers) from the protein. Therefore, measuring the protein mass under these conditions provides information only on the protein conformation and stability, rather than composition of noncovalent complexes formed by the protein and its ligands. In addition to characterizing protein conformation and stability globally, the protein can be digested with an acidic protease (e.g., pepsin) under the slow exchange conditions, and MS (alone or in combination with LC separation) can be used to measure the deuterium content of each proteolytic fragment. This produces information on protein conformation and dynamics at the local level. A typical workflow diagram of an HDX MS experiment is shown in Figure 4.16.

Spatial resolution offered by HDX MS is usually limited only by the extent of proteolysis, which (along with other sample-handling steps) must be performed relatively quickly after quenching to avoid occurrence of significant *back-exchange* prior to MS measurements of the deuterium content of individual peptide fragments. In general, a large number of fragments, particularly overlapping ones, would

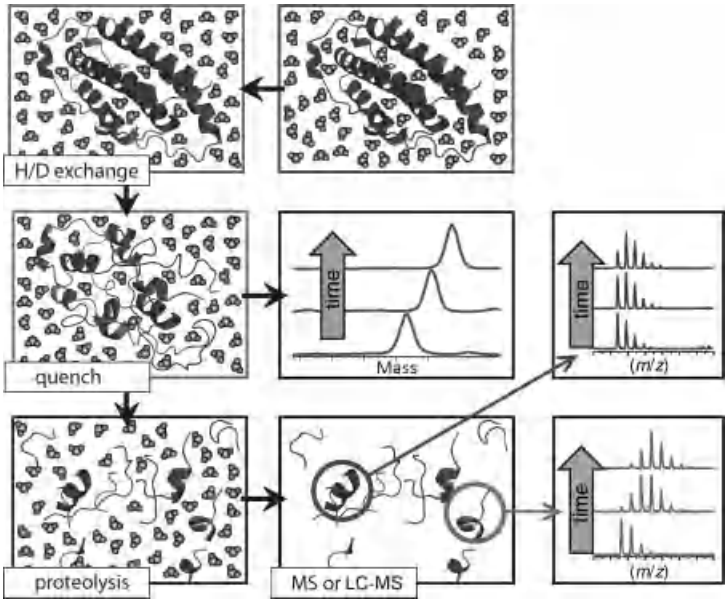


Figure 4.16. Schematic representation of HDX MS work flow to examine protein higher order structure and conformational dynamics. The exchange is initiated by placing the unlabeled protein into a D_2O–based solvent system (e.g., by a rapid dilution). Unstructured and highly dynamic protein segments undergo fast exchange (blue and red colors represent protons and deuterons, respectively). Following the quench step (rapid solution acidification and temperature drop), the protein loses its native conformation, but the spatial distribution of backbone amide protons and deuterons across the backbone is preserved (all labile hydrogen atoms at side chains undergo fast back-exchange at this step). Rapid clean-up followed by MS measurement of the protein mass reports the total number of backbone amide hydrogen atoms exchanged under native conditions (a global measure of the protein stability under native conditions), as long as the quench conditions are maintained during the sample work-up and measurement. Alternatively, the protein can be digested under the quench conditions using acid-stable protease(s), and LC–MS analysis of masses of individual proteolytic fragments will provide information on the backbone protection of corresponding protein segments under the native conditions. [Reprinted with permission from (182). Copyright © 2011 Bentham Science Publishers.] (See color version of the figure in Color Plate section)

lead to greater spatial resolution, and hence more precise localization of the structural regions that have undergone exchange. In some cases, this may allow the backbone amide protection patterns to be determined at single-residue resolution (183), although such instances are very rare. Supplementation of enzymatic digestion with peptide ion fragmentation in the gas phase may also enhance the spatial resolution of HDX MS measurements (184), but this technique has yet to be commonly accepted due to concerns over the possibility of introducing gas-phase artifacts (185). (see Chapter 5 for a more detailed discussion). In addition to limited spatial resolution, HDX MS measurements frequently suffer from incomplete sequence coverage, especially when applied to larger and extensively glycosylated proteins. Proteins with multiple disulfide bonds constitute another class of targets for which adequate sequence coverage is difficult to achieve, although certain changes in experimental protocol can alleviate this problem, at least for smaller proteins (186). Typically, an 80% level of sequence coverage is considered good, although significantly lower levels may also be adequate, depending on the context of the study.

Back-exchange, mentioned in the previous paragraph, is another pitfall of HDX MS measurements. Unfortunately, while undesirable, this is unavoidable even under controlled conditions, and in fact may be accelerated during various stages of protein processing, for example, during chromatographic separation of peptides (187). Although back-exchange was frequently evaluated in the past using unstructured model peptides, the utility of this procedure is questionable, since the intrinsic exchange rates are highly sequence-dependent. In many instances, back-exchange may be estimated using algorithms based on context-specific kinetics data (e.g., available at http://hx2.med.upenn.edu/download.html); it may also be determined experimentally for each proteolytic fragment by processing a fully labeled protein using a series of steps that precisely reproduce those used in HDX MS measurements (148). Typical back-exchange levels reported in the recent literature range from 10 to 50%, although significantly higher numbers have

also been reported. Even if back-exchange can be accounted for, it nonetheless has a very detrimental influence on the quality of HDX MS measurements by reducing the available dynamic range. Unlike chemical cross-linking and oxidative labeling, data analysis itself does not appear to be a significant problem in HDX MS work. Once all relevant peptic fragments are identified and back-exchange levels measured for all of them, calculation of local protection levels is relatively straightforward. Nonetheless, this task is very time consuming and can be streamlined by automated HDX MS data processing algorithms, especially for large proteins and in situations where multiple exchange data points are acquired (188–190).

An example of using HDX MS to identify interface regions in protein complexes is shown in Figure 4.17, where hydrogen-exchange kinetics are measured for a diferric form of human serum transferrin (Fe_2Tf) alone and when complexed to its cognate receptor. Both Tf–metal and Tf–receptor complexes dissociate under the slow exchange conditions prior to MS analysis; therefore, the protein mass evolution in each case reflects solely deuterium uptake in the course of exchange in solution. The extra protection afforded by the receptor binding to Tf persists over an extended period of time, and it may be tempting to assign it to shielding of labile hydrogen atoms at the protein–receptor interface. However, this view is overly simplistic, as the conformational effects of protein binding are frequently felt well beyond the interface region. The difference in the backbone protection levels of receptor-free and receptor-bound forms of Fe_2Tf appears to grow during the initial hour of exchange (Fig. 4.17a), reflecting significant stabilization of Fe_2Tf higher order structure by the bound receptor. Indeed, while the fast phase of HDX is typically ascribed to frequent local fluctuations (transient perturbations of higher order structure) affecting relatively small protein segments, the slower phases of HDX usually reflect relatively rare, large-scale conformational transitions (transient partial or complete unfolding). This explains why global HDX MS

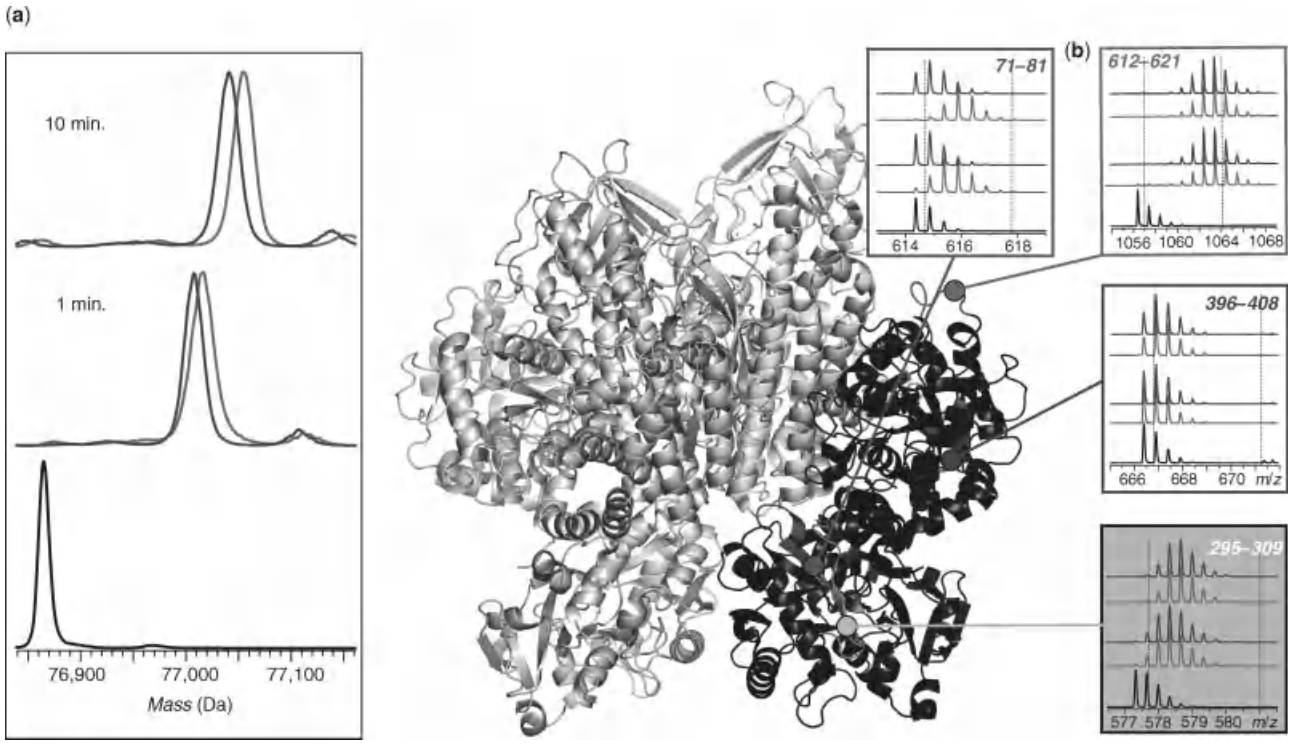


Figure 4.17. Localization of the receptor binding interface on the surface of human serum transferrin (Tf) with HDX MS. (*a*) The HDX MS of Tf (global exchange) in the presence (blue) and the absence (red) of the receptor. The exchange was carried out by diluting the protein stock solution 1:10 in exchange solution (100 mM NH_4HCO_3 in D_2O, pH adjusted to 7.4) and incubating for a certain period of time as indicated on each diagram followed by rapid quenching (lowering pH to 2.5 and temperature to near 0°C). The black trace shows unlabeled protein. (*b*) Isotopic distributions of representative peptic fragments derived from Tf subjected to HDX in the presence (blue) and the absence (red) of the receptor and followed by rapid quenching, proteolysis and LC MS analysis. Dotted lines indicate average masses of unlabeled and fully exchanged peptide ions. Colored segments within the Tf–receptor complex show localization of these peptic fragments [based on the low-resolution structure of the complex (190)]. [Adapted with permission from (185). Copyright © 2009 American Chemical Society.] (See color version of the figure in Color Plate section)

measurements similar to those presented in Figure 4.17 can be used to obtain quantitative thermodynamic characteristics for protein interaction with a variety of ligands, which will be considered in greater detail in Chapter 7. Evolution of deuterium content of various peptic fragments of Fe_2Tf (Fig. 4.17b) reveals a wide spectrum of protection, which is distributed very unevenly across the protein sequence. While some peptides exhibit nearly complete protection of backbone amides (e.g., segment [396–408] sequestered in the core of the protein C-lobe), exchange in many others is fast (e.g., peptide [612–621] in the solvent-exposed loop of the C-lobe). The influence of receptor binding on backbone protection is also highly localized. While most segments appear to be unaffected by the receptor binding, there are several regions where exchange kinetics are noticeably decelerated (e.g., segment [71–81] of the N-lobe, which contains several amino acid residues that form Tf–receptor interface according to the available model of the complex based on low-resolution cryo-EM data (191)).

Although HDX MS measurements alone are usually insufficient to determine protein conformation or quaternary structure of multiunit assemblies, combining it with other biophysical tools may lead to significant advances in structure characterization. This was demonstrated recently by Komives and co-workers (192), who used a combination of HDX MS and crystallography to establish differences and similarities among various forms of a viral capsid. The HDX MS data can also aid in building high-resolution quaternary structures of protein assemblies from subunits whose monomeric structures are known, by providing a filter for a vast array of suboptimal candidate structures produced by computational docking (193). Finally, HDX MS has been extremely helpful in mapping intersubunit interfaces in large protein aggregates and amyloid formations (183,194–196), which are notoriously difficult to analyze using traditional experimental approaches. These studies will be discussed in more detail in Chapter 9.

Although the focus of this chapter is on characterizing native protein structures, the experimental methods presented here can be used to probe dynamic features of biomolecules as well. In the following chapters, we will discuss applications of both chemical cross-linking and HDX to characterize the structure of various non-native states of proteins that become populated under denaturing conditions (Chapter 5). We will also consider various uses of HDX MS and oxidative chemical labeling to probe the transient protein states that play important roles in folding processes (Chapter 6).

REFERENCES

1. Mattos, C. (2002) Protein–water interactions in a dynamic world, *Trends Biochem. Sci. 27*, 203–208.
2. Breuker, K., Oh, H. B., Horn, D. M., Cerda, B. A., and McLafferty, F. W. (2002) Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation, *J. Am. Chem. Soc. 124*, 6407–6420.
3. Oh, H., Breuker, K., Sze, S. K., Ge, Y., Carpenter, B. K., and McLafferty, F. W. (2002) Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy, *Proc. Natl. Acad. Sci. USA. 99*, 15863–15868.
4. Weinkauf, R., Schermann, J. P., de Vries, M. S., and Kleinermanns, K. (2002) Molecular physics of building blocks of life under isolated or defined conditions, *Eur. Phys. J. D 20*, 309–316.
5. Hoaglund-Hyzer, C. S., Counterman, A. E., and Clemmer, D. E. (1999) Anhydrous protein ions, *Chem. Rev. 99*, 3037–3080.
6. Jarrold, M. F. (2000) Peptides and proteins in the vapor phase, *Annu. Rev. Phys. Chem. 51*, 179–207.
7. Barylyuk, K., Balabin, R., Grünstein, D., Kikkeri, R., Frankevich, V., Seeberger, P., and Zenobi, R. (2011) What happens to hydrophobic interactions during transfer from the solution to the gas phase? The case of electrospray-based soft ionization methods, *J. Am. Soc. Mass Spectrom. 22*, 1167–1177.
8. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev. 16*, 1–23.
9. Hernandez, H. and Robinson, C. V. (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry, *Nat. Protoc. 2*, 715–726.
10. Yin, S. and Loo, J. A. (2009) Mass spectrometry detection and characterization of noncovalent protein complexes, *Methods Mol. Biol. 492*, 273–282.
11. Heck, A. J. R. (2008) Native mass spectrometry: a bridge between interactomics and structural biology, *Nat. Meth. 5*, 927–933.
12. Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J. L., Wingfield, P. T., Steven, A. C., and Heck, A. J. R. (2008) High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids, *Proc. Natl. Acad. Sci. USA 105*, 9216–9220.
13. Abzalimov, R. R., Frimpong, A. K., and Kaltashov, I. A. (2006) Gas-phase processes and measurements of macromolecular properties in solution: On the possibility of false positive and false negative signals of protein unfolding, *Int. J. Mass Spectrom. 253*, 207–216.
14. Freeke, J., Robinson, C. V., and Ruotolo, B. T. (2009) Residual counter ions can stabilise a large protein complex in the gas phase, *Int. J. Mass Spectrom. 298*, 91–98.
15. McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes, *J. Am. Chem. Soc. 128*, 11433–11442.
16. Liepold, L., Oltrogge, L. M., Suci, P. A., Young, M. J., and Douglas, T. (2009) Correct charge state assignment of native electrospray spectra of protein complexes, *J. Am. Soc. Mass Spectrom. 20*, 435–442.

17. Tseng, Y.-H., Uetrecht, C., Heck, A. J. R., and Peng, W.-P. (2011) Interpreting the charge state assignment in electrospray mass spectra of bioparticles, *Anal. Chem. 83*, 1960–1968.
18. Abzalimov, R. R. and Kaltashov, I. A. (2010) Electrospray ionization mass spectrometry of highly heterogeneous protein systems: Protein ion charge state assignment via incomplete charge reduction, *Anal. Chem. 82*, 7523–7526.
19. Lei, Q. P., Cui, X., Kurtz, D. M., Jr., Amster, I. J., Chernushevich, I. V., and Standing, K. G. (1998) Electrospray mass spectrometry studies of non-heme iron-containing proteins, *Anal. Chem. 70*, 1838–1846.
20. Youn, H. D., Kim, E. J., Roe, J. H., Hah, Y. C., and Kang, S. O. (1996) A novel nickel-containing superoxide dismutase from Streptomyces spp, *Biochem. J. 318*, 889–896.
21. Wuerges, J., Lee, J.-W., Yim, Y.-I., Yim, H.-S., Kang, S.-O., and Carugo, K. D. (2004) Crystal structure of nickel-containing superoxide dismutase reveals another type of active site, *Proc. Natl. Acad. Sci. USA 101*, 8569–8574.
22. Rostom, A. A. and Robinson, C. V. (1999) Disassembly of intact multiprotein complexes in the gas phase, *Curr. Opin. Struct. Biol. 9*, 135–141.
23. Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M., and Robinson, C. V. (2000) Detection and selective dissociation of intact ribosomes in a mass spectrometer, *Proc. Natl. Acad. Sci. USA. 97*, 5185–5190.
24. Tito, M. A., Miller, J., Walker, N., Griffin, K. F., Williamson, E. D., Despeyroux-Hill, D., Titball, R. W., and Robinson, C. V. (2001) Probing molecular interactions in intact antibody: antigen complexes, an electrospray time-of-flight mass spectrometry approach, *Biophys. J. 81*, 3503–3509.
25. Benesch, J. L. P., Aquilina, J. A., Ruotolo, B. T., Sobott, F., and Robinson, C. V. (2006) Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies, *Chem. Biol. 13*, 597–605.
26. Tito, M. A., Miller, J., Griffin, K. F., Williamson, E. D., Titball, R. W., and Robinson, C. V. (2001) Macromolecular organization of the Yersinia pestis capsular F1 antigen: insights from time-of-flight mass spectrometry, *Protein Sci. 10*, 2408–2413.
27. Pinkse, M. W., Maier, C. S., Kim, J. I., Oh, B. H., and Heck, A. J. R. (2003) Macromolecular assembly of *Helicobacter pylori* urease investigated by mass spectrometry, *J. Mass Spectrom. 38*, 315–320.
28. Jurchen, J. C. and Williams, E. R. (2003) Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers, *J. Am. Chem. Soc. 125*, 2817–2826.
29. Jurchen, J. C., Garcia, D. E., and Williams, E. R. (2003) Gas-phase dissociation pathways of multiply charged peptide clusters, *J. Am. Soc. Mass Spectrom. 14*, 1373–1386.
30. Kaltashov, I. A. and Abzalimov, R. R. (2008) Do ionic charges in ESI MS provide useful information on macromolecular structure? *J. Am. Soc. Mass Spectrom. 19*, 1239–1246.
31. de la Mora, J. F. (1996) On the outcome of the Coulombic fission of a charged isolated drop, *J. Coll. Int. Sci. 178*, 209–218.
32. Chowdhury, S. K., Katta, V., and Chait, B. T. (1990) Probing conformational changes in proteins by mass spectrometry, *J. Am. Chem. Soc. 112*, 9012–9013.
33. Loo, J. A., Loo, R. R., Udseth, H. R., Edmonds, C. G., and Smith, R. D. (1991) Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 5*, 101–105.
34. Fenselau, C., Szilagyi, Z., and Williams, T. (2000) Intercharge distances in Zn7-metallothionein analyzed by nanospray on a quadrupole ion trap and molecular modeling, *J. Mass Spectrom. Soc. Jpn. 48*, 23–25.
35. de la Mora, J. F. (2000) Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Analyt. Chim. Acta 406*, 93–104.
36. Heck, A. J. R. and van Den Heuvel, R. H. (2004) Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom. Rev. 23*, 368–389.
37. Nesatyy, V. J. and Suter, M. J. F. (2004) On the conformation-dependent neutralization theory and charging of individual proteins and their non-covalent complexes in the gas phase, *J. Mass Spectrom. 39*, 93–97.
38. Kaltashov, I. A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal. Chem. 77*, 5370–5379.
39. Mendoza, V. L., Antwi, K., Barón-Rodríguez, M. A., Blanco, C., and Vachet, R. W. (2010) Structure of the preamyloid dimer of β-2-microglobulin from covalent labeling and mass spectrometry, *Biochemistry 49*, 1522–1532.
40. Rayleigh, J. W. S. (1882) On the equilibrium of liquid conducting masses charged with electricity, *Philos. Mag. 14*, 184–186.
41. de Juan, L. and de la Mora, J. F. (1997) Charge and size distributions of electrospray drops, *J. Colloid Interf. Sci. 186*, 280–293.
42. Hartman, R. P. A., Brunner, D. J., Camelot, D. M. A., Marijnissen, J. C. M., and Scarlett, B. (2000) Jet break-up in electrohydrodynamic atomization in the cone-jet mode, *J. Aerosol Sci. 31*, 65–95.
43. Liu, J. and Konermann, L. (2011) Protein-protein binding affinities in solution determined by electrospray mass spectrometry, *J. Am. Soc. Mass Spectrom. 22*, 408–417.
44. Testa, L., Brocca, S., and Grandori, R. (2011) Charge-surface correlation in electrospray ionization of folded and unfolded proteins, *Anal. Chem. 83*, 6459–6463.
45. Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E., and Heck, A. J. R. (2011) Interrogating viral capsid assembly with ion mobility/mass spectrometry, *Nat. Chem. 3*, 126–132.
46. Pukala, T. L., Ruotolo, B. T., Zhou, M., Politis, A., Stefanescu, R., Leary, J. A., and Robinson, C. V. (2009) Subunit architecture of multiprotein assemblies determined using restraints from gas-phase measurements, *Structure 17*, 1235–1243.
47. Breuker, K. and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^{2} s, *Proc. Natl. Acad. Sci. USA 105*, 18145–18152.

48. Jurneczko, E. and Barran, P. E. (2011) How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase, *Analyst 136*, 20–28.
49. Hogan, C. J., Ruotolo, B. T., Robinson, C. V., and Fernandez de la Mora, J. (2011) Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the GroEL complex, *J. Phys. Chem. B 115*, 3614–3621.
50. Michaelevski, I., Eisenstein, M., and Sharon, M. (2010) Gas-phase compaction and unfolding of protein structures, *Anal. Chem. 82*, 9484–9491.
51. Politis, A., Park, A. Y., Hyung, S.-J., Barsky, D., Ruotolo, B. T., and Robinson, C. V. (2010) Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes, *PLoS ONE 5*, e12080.
52. Kaddis, C., Lomeli, S., Yin, S., Berhane, B., Apostol, M., Kickhoefer, V., Rome, L., and Loo, J. (2007) Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility, *J. Am. Soc. Mass Spectrom. 18*, 1206–1216.
53. Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L., and Chernushevich, I. V. (2005) Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex, *J. Am. Soc. Mass Spectrom. 16*, 998–1008.
54. Laschober, C., Wruss, J., Blaas, D., Szymanski, W. W., and Allmaier, G. (2008) Gas-phase electrophoretic molecular mobility analysis of size and stoichiometry of complexes of a common cold virus with antibody and soluble receptor molecules, *Anal. Chem. 80*, 2261–2264.
55. Kapellios, E., Karamanou, S., Sardis, M., Aivaliotis, M., Economou, A., and Pergantis, S. (2011) Using nanoelectrospray ion mobility spectrometry (GEMMA) to determine the size and relative molecular mass of proteins and protein assemblies: a comparison with MALLS and QELS, *Anal. Bioanal. Chem. 399*, 2421–2433.
56. Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K., and Heck, A. J. R. (2010) Ion mobility mass spectrometry of proteins and protein assemblies, *Chem. Soc. Rev. 39*, 1633–1655.
57. Juraschek, R., Dulcks, T., and Karas, M. (1999) Nanoelectrospray—more than just a minimized-flow electrospray ionization source, *J. Am. Soc. Mass Spectrom. 10*, 300–308.
58. Iavarone, A. T., Udekwu, O. A., and Williams, E. R. (2004) Buffer loading for counteracting metal salt-induced signal suppression in electrospray ionization, *Anal. Chem. 76*, 3944–3950.
59. Sterling, H. J., Batchelor, J. D., Wemmer, D. E., and Williams, E. R. (2010) Effects of buffer loading for electrospray ionization mass spectrometry of a noncovalent protein complex that requires high concentrations of essential salts, *J. Am. Soc. Mass. Spectrom. 21*, 1045–1049.
60. Kaltashov, I. A., Abzalimov, R. R., Eyles, S. J., and Frimpong, A., (2008) Studies of intact proteins and protein complexes. ESI MS approaches, In *Mass Spectrometry Analyses for Protein–Protein Interactions and Dynamics* (Chance, M. R. Ed.) pp 215–240 Wiley Blackwell, Boston.
61. Glocker, M. O., Bauer, S. H. J., Kast, J., Volz, J., and Przybylski, M. (1996) Characterization of specific noncovalent protein complexes by UV matrix-assisted laser desorption ionization mass spectrometry, *J. Mass Spectrom. 31*, 1221–1227.
62. Moniatte, M., van der Goot, F. G., Buckley, J. T., Pattus, F., and van Dorsselaer, A. (1996) Characterisation of the heptameric pore-forming complex of the Aeromonas toxin aerolysin using MALDI–TOF mass spectrometry, *FEBS Lett. 384*, 269–272.
63. Moniatte, M., Lesieur, C., Vecsey-Semjen, B., Buckley, J. T., Pattus, F., van der Goot, F. G., and Van Dorsselaer, A. (1997) Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry in the subunit stoichiometry study of high-mass non-covalent complexes, *Int. J. Mass Spectrom. Ion Proc. 169–170*, 179–199.
64. Cohen, L. R. H., Strupat, K., and Hillenkamp, F. (1997) Analysis of quaternary protein ensembles by matrix assisted laser desorption/ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 8*, 1046–1052.
65. Lin, S. H., Cotter, R. J., and Woods, A. S. (1998) Detection of non-covalent interaction of single and double stranded DNA with peptides by MALDI–TOF, *Proteins*, 12–21.
66. Farmer, T. B. and Caprioli, R. M. (1998) Determination of protein-protein interactions by matrix-assisted laser desorption/ionization mass spectrometry, *J. Mass Spectrom. 33*, 697–704.
67. Kiselar, J. G. and Downard, K. M. (2000) Preservation and detection of specific antibody-peptide complexes by matrix-assisted laser desorption ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 11*, 746–750.
68. Woods, A. S. and Huestis, M. A. (2001) A study of peptide-peptide interaction by matrix-assisted laser desorption/ionization, *J. Am. Soc. Mass Spectrom. 12*, 88–96.
69. Zehl, M. and Allmaier, G. (2003) Investigation of sample preparation and instrumental parameters in the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of noncovalent peptide/peptide complexes, *Rapid Commun. Mass Spectrom. 17*, 1931–1940.
70. Pimenova, T., Pereira, C. P., Schaer, D.J., and Zenobi, R. (2009) Characterization of high molecular weight multimeric states of human haptoglobin and hemoglobin-based oxygen carriers by high-mass MALDI MS, *J. Sep. Sci. 32*, 1224–1230.
71. Bolbach, G. (2005) Matrix-assisted laser desorption/ionization analysis of non-covalent complexes: Fundamentals and applications, *Curr. Pharm. Des. 11*, 2535–2557.
72. Zahn, H. and Meienhofer, J. (1958) Reacktionen von 1,5-difluor-2,4-dinitrobenzol mit insulin. 1. Synthese von modellverbindungen, *Makromol. Chem. 26*, 126–152.
73. Zahn, H. and Meienhofer, J. (1958) Reaktionen von 1,5-difluor-2,4-dinitrobenzol mit insulin. 2. Mitt. versuche mit insulin, *Makromol. Chem. 26*, 153–166.
74. Hvidt, A. and Linderstrøm-Lang, K. (1954) Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions, *Biochim. Biophys. Acta 14*, 574–575.

75. Hvidt, A. and Linderstrøm-Lang, K. (1955) The kinetics of deuterium exchange of insulin with D_2O. An amendment, *Biochim. Biophys. Acta 16*, 168–169.
76. Wong, S. S. (1991) *Chemistry of protein conjugation and cross-linking*, CRC Press, Boca Raton, FL.
77. Green, N. S., Reisler, E., and Houk, K. N. (2001) Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers, *Protein Sci. 10*, 1293–1304.
78. Han, K.-K., Richard, C., and Delacourte, A. (1984) Chemical cross-links of proteins by using bifunctional reagents, *Int. J. Biochem. 16*, 129–145.
79. Mattson, G., Conklin, E., Desai, S., Nielander, G., Savage, M. D., and Morgensen, S. (1993) A practical approach to crosslinking, *Mol. Biol. Rep. 17*, 167–183.
80. Andrea, S. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions, *Mass Spectrom. Rev. 25*, 663–682.
81. Petrotchenko, E. V. and Borchers, C. H. (2010) Crosslinking combined with mass spectrometry for structural proteomics, *Mass Spectrom. Rev. 29*, 862–876.
82. Potluri, S., Khan, A. A., Kuzminykh, A., Bujnicki, J. M., Friedman, A. M., and Bailey-Kellogg, C. (2004) Geometric analysis of cross-linkability for protein fold discrimination, *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing 2004*, 447–458.
83. Zelter, A., Hoopmann, M. R., Vernon, R., Baker, D., MacCoss, M. J., and Davis, T. N. (2010) Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: A novel method to determine interresidue distances in protein structures through cross-linking, *J. Proteome Res. 9*, 3583–3589.
84. Kahraman, A., Malmstrm, L., and Aebersold, R. (2011) Xwalk: Computing and visualizing distances in cross-linking experiments, *Bioinformatics 27*, 2163–2164.
85. Rappsilber, J., Siniossoglou, S., Hurt, E. C., and Mann, M. (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry, *Anal. Chem. 72*, 267–275.
86. Back, J. W., de Jong, L., Muijsers, A. O., and de Koster, C. G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling, *J. Mol. Biol. 331*, 303–313.
87. Jin Lee, Y. (2008) Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes, *Mol. Biosyst. 4*, 816–823.
88. Bossi, A., Patel, M. J., Webb, E. J., Baldwin, M. A., Jacob, R. J., Burlingame, A. L., and Righetti, P. G. (1999) Analysis of cross-linked human hemoglobin by conventional isoelectric focusing, immobilized pH gradients, capillary electrophoresis, and mass spectrometry, *Electrophoresis 20*, 2810–2817.
89. Yang, T., Horejsh, D. R., Mahan, K. J., Zaluzec, E. J., Watson, T. J., and Gage, D. A. (1996) Mapping cross-linking sites in modified proteins with mass spectrometry: an application to cross-linked hemoglobins, *Anal. Biochem. 242*, 55–63.
90. Bennett, K. L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000) Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping—a novel approach to assess intermolecular protein contacts, *Protein Sci. 9*, 1503–1518.
91. Sinz, A. and Wang, K. (2001) Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin–calmodulin complexes, *Biochemistry 40*, 7903–7913.
92. Yu, Z., Friso, G., Miranda, J. J., Patel, M. J., Lo-Tseng, T., Moore, E. G., and Burlingame, A. L. (1997) Structural characterization of human hemoglobin crosslinked by bis (3,5-dibromosalicyl) fumarate using mass spectrometric techniques, *Protein Sci. 6*, 2568–2577.
93. Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry, *Proc. Natl. Acad. Sci. USA 97*, 5802–5806.
94. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, *J. Am. Soc. Mass Spectrom. 14*, 834–850.
95. Chu, F., Mahrus, S., Craik, C. S., and Burlingame, A. L. (2006) Isotope-coded and affinity-tagged cross-linking (ICATXL): An efficient strategy to probe protein interaction surfaces, *J. Am. Chem. Soc. 128*, 10362–10363.
96. Chen, X., Chen, Y. H., and Anderson, V. E. (1999) Protein cross-links: universal isolation and characterization by isotopic derivatization and electrospray ionization mass spectrometry, *Anal. Biochem. 273*, 192–203.
97. Müller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis, *Anal. Chem. 73*, 1927–1934.
98. Back, J. W., Hartog, A. F., Dekker, H. L., Muijsers, A. O., de Koning, L. J., and de Jong, L. (2001) A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes, *J. Am. Soc. Mass Spectrom. 12*, 222–227.
99. Dihazi, G. H. and Sinz, A. (2003) Mapping low-resolution three-dimensional protein structures using chemical cross-linking and Fourier transform ion-cyclotron resonance mass spectrometry, *Rapid Commun. Mass Spectrom. 17*, 2005–2014.
100. Person, M. D., Brown, K. C., Mahrus, S., Craik, C. S., and Burlingame, A. L. (2001) Novel inter-protein cross-link identified in the GGH-ecotin D137Y dimer, *Protein Sci. 10*, 1549–1562.
101. Trnka, M. J. and Burlingame, A. L. (2010) Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene, *Mol. Cell. Proteometrics. 9*, 2306–2317.
102. Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010) Probing native protein structures by chemical cross-linking, mass

spectrometry, and bioinformatics, *Mol. Cell. Proteomics 9*, 1634–1649.

103. Winters, M. S. and Day, R. A. (2003) Detecting protein-protein interactions in the intact cell of Bacillus subtilis (ATCC 6633), *J. Bacteriol. 185*, 4268–4275.
104. Du, X., Chowdhury, S. M., Manes, N. P., Wu, S., Mayer, M. U., Adkins, J. N., Anderson, G. A., and Smith, R. D. (2011) XlinkIdentifier: An automated data analysis platform for confident identifications of chemically cross-linked peptides using tandem mass spectrometry, *J. Proteome Res. 10*, 923–931.
105. Chu, F., Baker, P. R., Burlingame, A. L., and Chalkley, R. J. (2010) Finding chimeras: a bioinformatics strategy for identification of cross-linked peptides, *Mol. Cell. Proteomies. 9*, 25–31.
106. Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008) Identification of cross-linked peptides from large sequence databases, *Nature Met. 5*, 315–318.
107. Gao, Q. X., Xue, S., Doneanu, C. E., Shaffer, S. A., Goodlett, D. R., and Nelson, S. D. (2006) Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry, *Anal. Chem. 78*, 2145–2149.
108. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007) Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching, *Mol. Cell. Proteomis. 6*, 2200–2211.
109. Petrotchenko, E. V. and Borchers, C. H. (2010) ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite, *BMC Bioinformatics 11*, 64–73.
110. Jacob, C., Giles, G. I., Giles, N. M., and Sies, H. (2003) Sulfur and selenium: the role of oxidation state in protein structure and function, *Angew. Chem. Int. Ed. Engl. 42*, 4742–4758.
111. van Vlijmen, H. W. T., Gupta, A., Narasimhan, L. S., and Singh, J. (2004) A novel database of disulfide patterns and its application to the discovery of distantly related homologs, *J. Mol. Biol. 335*, 1083–1092.
112. Thangudu, R. R., Vinayagam, A., Pugalenthi, G., Manonmani, A., Offmann, B., and Sowdhamini, R. (2005) Native and modeled disulfide bonds in proteins: Knowledge-based approaches toward structure prediction of disulfide-rich polypeptides, *Proteins 58*, 866–879.
113. Dillon, T. M., Bondarenko, P. V., Rehder, D. S., Pipes, G. D., Kleemann, G. R., and Ricci, M. S. (2006) Optimization of a reversed-phase high-performance liquid chromatography/ mass spectrometry method for characterizing recombinant antibody heterogeneity and stability, *J. Chromatogr. A 1120*, 112–120.
114. Bagal, D., Valliere-Douglass, J. F., Balland, A., and Schnier, P. D. (2010) Resolving disulfide structural isoforms of IgG2 monoclonal antibodies by ion mobility mass spectrometry, *Anal. Chem. 82*, 6751–6755.
115. Bilusich, D., Maselli, V. M., Brinkworth, C. S., Samguina, T., Lebedev, A. T., and Bowie, J. H. (2005) Direct identification of intramolecular disulfide links in peptides using negative ion electrospray mass spectra of underivatised peptides. A joint experimental and theoretical study, *Rapid Commun. Mass Spectrom. 19*, 3063–3074.
116. Bilusich, D. and Bowie, J. H. (2007) Identification of intermolecular disulfide linkages in underivatised peptides using negative ion electrospray mass spectrometry. A joint experimental and theoretical study, *Rapid Commun. Mass Spectrom. 21*, 619–628.
117. Chrisman, P. A. and McLuckey, S. A. (2002) Dissociations of disulfide-linked gaseous polypeptide/protein anions: ion chemistry with implications for protein identification and characterization, *J. Proteome Res. 1*, 549–557.
118. Zhang, M. and Kaltashov, I. A. (2006) Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection, *Anal. Chem. 78*, 4820–4829.
119. Nielsen, R. W., Tachibana, C., Hansen, N. E., and Winther, J. R. (2011) Trisulfides in Proteins, *Antioxidants Redox Signaling 15*, 67–75.
120. Gu, S., Wen, D., Weinreb, P. H., Sun, Y., Zhang, L., Foley, S. F., Kshirsagar, R., Evans, D., Mi, S., Meier, W., and Pepinsky, R. B. (2010) Characterization of trisulfide modification in antibodies, *Anal. Biochem. 400*, 89–98.
121. Pristatsky, P., Cohen, S. L., Krantz, D., Acevedo, J., Ionescu, R., and Vlasak, J. (2009) Evidence for trisulfide bonds in a recombinant variant of a human IgG2 monoclonal antibody, *Anal. Chem. 81*, 6148–6155.
122. Qiu, X. Y., Cui, M., Liu, Z. Q., and Liu, S. Y. (2008) Protein disulfide bond determination and its analysis by mass spectrometry, *Progr. Chem. 20*, 975–983.
123. Endrizzi, B. J., Huang, G., Kiser, P. F., and Stewart, R. J. (2006) Specific covalent immobilization of proteins through dityrosine cross-links, *Langmuir 22*, 11305–11310.
124. Person, M. D., Brown, K. C. Mahues, S., Craik, C. S., and Burlingame, A. L. (2001) Novel inter-protein cross-link identified in the GGH-ecotin D137Y dimer, *Protein Sci. 10*, 1549–1562
125. Brown, K. C., Yu, Z., Burlingame, A. L., and Craik, C. S. (1998) Determining protein-protein interactions by oxidative cross-linking of a glycine-glycine-histidine fusion protein, *Biochemistry 37*, 4397–4406.
126. Meunier, S., Strable, E., and Finn, M. G. (2004) Crosslinking of and coupling to viral capsid proteins by tyrosine oxidation, *Chem. Biol. 11*, 319–326.
127. Glazer, A. N. (1970) Specific Chemical Modification of Proteins, *Annu. Rev. Biochem. 39*, 101–130.
128. Gundlach, H. G., Stein, W. H., and Moore, S. (1959) The nature of the amino acid residues involved in the inactivation of ribonuclease by iodoacetate, *J. Biol. Chem. 234*, 1754–1760.
129. Shaw, D. C., Stein, W. H., and Moore, S. (1964) Inactivation of chymotrypsin by cyanate, *J. Biol. Chem. 239*, 671–673.
130. Rajagopalan, T. G., Stein, W. H., and Moore, S. (1966) The inactivation of pepsin by diazoacetylnorleucine methyl ester, *J. Biol. Chem. 241*, 4295–4297.
131. Jackson, A. E., Carraway, K. L., Puett, D., and Brew, K. (1986) Effects of the binding of myosin light chain kinase on

the reactivities of calmodulin lysines, *J. Biol. Chem. 261*, 12226–12232.

132. Steiner, R. F., Albaugh, S., Fenselau, C., Murphy C., and Vestling, M. (1991) A mass spectrometry method for mapping the interface topography of interacting proteins, illustrated by the melittin-calmodulin system, *Anal. Biochem. 196*, 120–125.
133. Mendoza, V. L. and Vachet, R. W. (2009) Probing protein structure by amino acid-specific covalent labeling and mass spectrometry, *Mass Spectrom. Rev. 28*, 785–815.
134. Hager-Braun, C. and Tomer, K. B. (2002) Characterization of the tertiary structure of soluble CD4 bound to glycosylated full-length HIVgp120 by chemical modification of arginine residues and mass spectrometric analysis, *Biochemistry 41*, 1759–1766.
135. Hochleitner, E. O., Borchers, C., Parker, C., Bienstock, R. J., and Tomer, K. B. (2000) Characterization of a discontinuous epitope of the human immunodeficiency virus (HIV) core protein p24 by epitope excision and differential chemical modification followed by mass spectrometric peptide mapping analysis, *Protein Sci. 9*, 487–496.
136. Happersberger, H. P., Przybylski, M., and Glocker, M. O. (1998) Selective bridging of bis-cysteinyl residues by arsonous acid derivatives as an approach to the characterization of protein tertiary structures and folding pathways by mass spectrometry, *Anal. Biochem. 264*, 237–250.
137. Zappacosta, F., Ingallinella, P., Scaloni, A., Pessi, A., Bianchi, E., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1997) Surface topology of Minibody by selective chemical modifications and mass spectrometry, *Protein Sci. 6*, 1901–1909.
138. Avaeva, S. M. and Krasnova, V. I. (1975) Diethylpyrocarbonate reaction with imidazole and histidine derivatives, *Bioorg. Khim. 1*, 1600–1605.
139. Riordan, J. F., Sokolovsky, M., and Vallee, B. L. (1967) Environmentally sensitive tyrosyl residues. Nitration with tetranitromethane, *Biochemistry 6*, 358–361.
140. Sokolovsky, M., Riordan, J. F., and Vallee, B. L. (1966) Tetranitromethane. A reagent for the nitration of tyrosyl residues in proteins, *Biochemistry 5*, 3582–3589.
141. Glazer, A. N. (1976) The chemical modification of proteins by group-specific and site-specific reagents, In *The proteins* (Neurath, H., Hill, R. L., and Boeder, C. L., Eds.) 3rd ed., pp 1–103, Academic Press, New York.
142. Liu, X. and Reilly, J. P. (2009) Correlating thechemical modification of *Escherichia coli* ribosomal proteins with crystal structure data, *J. Proteome Res. 8*, 4466–4478.
143. Toi, K., Bynum, E., Norris, E., and Itano, H. A. (1967) Studies on the chemical modification of arginine. I. The reaction of 1,2-cyclohexanedione with arginine and arginyl residues of proteins, *J. Biol. Chem. 242*, 1036–1043.
144. Toi, K., Bynum, E., Norris, E., and Itano, H. A. (1965) Chemical Modification of Arginine with 1,2-Cyclohexanedione, *J. Biol. Chem. 240*, 3455–3457.
145. Calvete, J. J., Campanero-Rhodes, M. A., Raida, M., and Sanz, L. (1999) Characterisation of the conformational and quaternary structure-dependent heparin-binding region of bovine seminal plasma protein PDC-109, *FEBS Lett. 444*, 260–264.
146. Suckau, D., Mak, M., and Przybylski, M. (1992) Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping, *Proc. Natl. Acad. Sci. USA 89*, 5630–5634.
147. Bobst, C. E., Abzalimov, R. R., Houde, D., Kloczewiak, M., Mhatre, R., Berkowitz, S. A., and Kaltashov, I. A. (2008) Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches, *Anal. Chem. 80*, 7473–7481.
148. Sharp, J. S., Becker, J. M., and Hettich, R. L. (2003) Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry, *Anal. Biochem. 313*, 216–225.
149. Lim, J. and Vachet, R. W. (2003) Development of a methodology based on metal-catalyzed oxidation reactions and mass spectrometry to determine the metal binding sites in copper metalloproteins, *Anal. Chem. 75*, 1164–1172.
150. Scaloni, A., Ferranti, P., De Simone, G., Mamone, G., Sannolo, N., and Malorni, A. (1999) Probing the reactivity of nucleophile residues in human 2,3-diphosphoglycerate/deoxy-hemoglobin complex by aspecific chemical modifications, *FEBS Lett. 452*, 190–194.
151. Maleknia, S. D., Chance, M. R., and Downard, K. M. (1999) Electrospray-assisted modification of proteins: a radical probe of protein structure, *Rapid Commun. Mass Spectrom. 13*, 2352–2358.
152. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than rrotein unfolding, *Anal. Chem. 81*, 6563–6571.
153. Goshe, M. B., Chen, Y. H., and Anderson, V. E. (2000) Identification of the sites of hydroxyl radical reaction with peptides by hydrogen/deuterium exchange: prevalence of reactions with the side chains, *Biochemistry 39*, 1761–1770.
154. Kiselar, J. G., Maleknia, S. D., Sullivan, M., Downard, K. M., and Chance, M. R. (2002) Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry, *Int. J. Radiat. Biol. 78*, 101–114.
155. Xu, G., Takamoto, K., and Chance, M. R. (2003) Radiolytic modification of basic amino Acid residues in peptides: probes for examining protein–protein interactions, *Anal. Chem. 75*, 6995–7007.
156. Xu, G. H., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem. Rev. 107*, 3514–3543.
157. Konermann, L., Stocks, B. B., Pan, Y., and Tong, X. (2010) Mass spectrometry combined with oxidative labeling for exploring protein structure and folding, *Mass Spectrom. Rev. 29*, 651–667.
158. Roeser, J., Bischoff, R., Bruins, A. P., and Permentier, H. P. (2010) Oxidative protein labeling in mass-spectrometry-based proteomics, *Anal. Bioanal. Chem. 397*, 3441–3455.
159. Kiselar, J. G. and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting, *J. Mass Spectrom. 45*, 1373–1382.

160. Kaur, P., Kiselar, J. G., and Chance, M. R. (2009) Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry, *Anal. Chem. 81*, 8141–8149.

161. Hambly, D. and Gross, M. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *J. Am. Soc. Mass. Spectrom. 16*, 2057–2063.

162. Richards, F., Lamed, R., Wynn, R., Patel, D., and Olack, G. (2000) Methylene as a possible universal footprinting reagent that will include hydrophobic surface areas: overview and feasibility: properties of diazirine as a precursor, *Protein Sci. 9*, 2506–2517.

163. Craig, P. O., Ureta, D. B., and Delfino, J. M. (2002) Probing protein conformation with a minimal photochemical reagent, *Protein Sci. 11*, 1353–1366.

164. Ureta, D. B., Craig, P. O., Gomez, G. E., and Delfino, J. M. (2007) Assessing native and non-native conformational states of a protein by methylene carbene labeling: The case of Bacillus licheniformis β-lactamase, *Biochemistry 46*, 14567–14577.

165. Gómez, G. E., Cauerhff, A., Craig, P. O., Goldbaum, F. A., and Delfino, J. M. (2006) Exploring protein interfaces with a general photochemical reagent, *Protein Sci. 15*, 744–752.

166. Jumper, C. C. and Schriemer, D. C. (2011) Mass spectrometry of paser-initiated carbene reactions for protein topographic analysis, *Anal. Chem. 83*, 2913–2920.

167. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) Hydrogen exchange: the modern legacy of Linderstrom-Lang, *Protein Sci. 6*, 1101–1109.

168. Urey, H. C., Brickedde, F. G., and Murphy, G. M. (1932) A hydrogen isotope of mass 2, *Phys. Rev. 39*, 164–165.

169. Lewis, G. N. and Macdonald, R. T. (1933) Some properties of pure H^2H_2O, *J. Am. Chem. Soc. 55*, 3057–3059.

170. Bonhoeffer, K. F. and Klar, R. (1934) Uber den austausch von schweren wasserstoffatomen zwischen wasser und organischen verbindungen, *Naturwissenschaften 22*, 45–45.

171. Champetier, G. and Viallard, R. (1938) Reaction d'echange de la cellulose et de l'eau lourde. Hydratation de la cellulose, *B. Soc. Chim. 33*, 1042–1048.

172. Burley, R. W., Nicholls, C. H., and Speakman, J. B. (1955) The crystalline/amorphous ratio of keratin fibres. Part II. The hydrogen-deuterium exchange reaction, *J. Text. Inst. 46*, T427–T432.

173. Sethi, S. K., Smith, D. L., and McCloskey, J. A. (1983) Determination of active hydrogen content by fast atom bombardment mass spectrometry following hydrogen-deuterium exchange, *Biochem. Biophys. Res. Commun. 112*, 126–131.

174. Katta, V. and Chait, B. T. (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 5*, 214–217.

175. Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) Identification of protein-protein interfaces by decreased amide proton solvent accessibility, *Proc. Natl. Acad. Sci. USA 95*, 14705–14710.

176. Ehring, H. (1999) Hydrogen exchange/electrospray ionization mass spectrometry studies of structural features of proteins and protein/protein interactions, *Anal. Biochem. 267*, 252–259.

177. Engen, J. R. (2009) Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS, *Anal. Chem. 81*, 7870–7875.

178. Konermann, L., Pan, J., and Liu, Y.-H. (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics, *Chem. Soc. Rev. 40*, 1224–1234.

179. Englander, S. W. (2006) Hydrogen exchange and mass spectrometry: A historical perspective, *J. Am. Soc. Mass Spectrom. 17*, 1481–1489.

180. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M., and Griffin, P. R. (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions, *Exp. Rev. Proteom. 8*, 43–59.

181. Dempsey, C. E. (2001) Hydrogen exchange in peptides and proteins using NMR-spectroscopy, *Progr. Nucl. Magn. Res. Spectrosc. 39*, 135–170.

182. Bobst, C. E. and Kaltashov, I. A. (2011) Advanced mass spectrometry-based methods for the analyses of conformational integrity of biopharmaceutical products, *Curr. Pharm. Biotechnol*, *12*, 1517–1529.

183. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L., Jr. (2005) Structure and properties of α-synuclein and other amyloids determined at the amino acid level, *Proc. Natl. Acad. Sci. USA. 102*, 15477–15482.

184. Rand, K. D., Zehl, M., Jensen, O. N., and Jorgensen, T. J. D. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry, *Anal. Chem. 81*, 5577–5584.

185. Kaltashov, I. A., Bobst, C. E., and Abzalimov, R. R. (2009) H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: Is there a need for a top-down approach?, *Anal. Chem. 81*, 7892–7899.

186. Burke, J. E., Karbarz, M. J., Deems, R. A., Li, S., Woods, V. L., and Dennis, E. A. (2008) Interaction of group IA phospholipase A2 with metal ions and phospholipid vesicles probed with deuterium exchange mass spectrometry, *Biochemistry 47*, 6451–6459.

187. Zhang, H. M., Bou-Assaf, G. M., Emmett, M. R., and Marshall, A. G. (2008) Fast reversed-phase liquid chromatography to reduce back exchange and increase throughput in H/D exchange monitored by FT–ICR mass spectrometry, *J. Am. Soc. Mass Spectrom 20*, 520–524.

188. Pascal, B. D., Chalmers, M. J., Busby, S. A., Mader, C. C., Southern, M. R., Tsinoremas, N. F., and Griffin, P. R. (2007) The Deuterator: Software for the determination of backbone amide deuterium levels from H/D exchange MS data, *BMC Bioinformatics 8*, 156.

189. Pascal, B., Chalmers, M., Busby, S., and Griffin, P. (2009) HD desktop: An integrated platform for the analysis and

visualization of H/D exchange data, *J. Am. Soc. Mass. Spectrom. 20*, 601–610.

190. Liu, S., Liu, L., Uzuner, U., Zhou, X., Gu, M., Shi, W., Zhang, Y., Dai, S., and Yuan, J. (2011) HDX-Analyzer: a novel package for statistical analysis of protein structure dynamics, *BMC Bioinformatics 12*, S43.
191. Cheng, Y., Zak, O., Aisen, P., Harrison, S. C., and Walz, T. (2004) Structure of the human transferrin receptor-transferrin complex, *Cell 116*, 565–576.
192. Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., Hendrix, R. W., Komives, E. A., and Johnson, J. E. (2009) An unexpected twist in viral capsid maturation, *Nature (London) 458*, 646–650.
193. Anand, G. S., Law, D., Mandell, J. G., Snead, A. N., Tsigelny, I., Taylor, S. S., Eyck, L. F. T., and Komives, E. A. (2003) Identification of the protein kinase A regulatory RI α-catalytic subunit interface by amide H/^{2}H exchange and protein docking, *Proc. Natl. Acad. Sci. USA 100*, 13264–13269.
194. Damo, S. M., Phillips, A. H., Young, A. L., Li, S., Woods, V. L., and Wemmer, D. E. (2010) Probing the conformation of a prion protein fibril with hydrogen exchange, *J. Biol. Chem. 285*, 32303–32311.
195. Zhang, A., Qi, W., Good, T. A., and Fernandez, E. J. (2009) Structural differences between Aβ(1-40) intermediate oligomers and fibrils elucidated by proteolytic fragmentation and hydrogen/deuterium exchange, *Biophys. J. 96*, 1091–1104.
196. Carulla, N., Zhou, M., Arimon, M., Gaira, M., Giralt, E., Robinson, C. V., and Dobson, C. M. (2009) Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation, *Proc. Natl. Acad. Sci. USA 106*, 7828–7833.

5

MASS SPECTROMETRY BASED APPROACHES TO STUDY BIOMOLECULAR DYNAMICS: EQUILIBRIUM INTERMEDIATES

In Chapter 4, we surveyed various mass spectrometry based approaches to study the higher order structure of proteins under native conditions. For many decades, such well-defined and highly organized structures were thought of as the most important (if not the only) determinants of protein function. Protein folding had been considered a linear process leading from fully unstructured (and, therefore, dysfunctional) states to the highly organized native (functionally competent) state. The advent of nuclear magnetic resonance changed our perception of what functional protein states are, with the realization that native proteins are very dynamic species. Perhaps the most illustrious examples of the intimate link between protein dynamics and function are found in enzyme catalysis, where the chemical conversion of substrate to product is often driven by relatively small-scale dynamic events within (and often beyond) the active site. It has become clear in recent years that large-scale macromolecular dynamics may also be an important determinant of protein function. A growing number of proteins are found to be either partially or fully unstructured under native conditions, and such flexibility (intrinsic disorder) appears to be vital for their function. Proteins that do have native folds under physiological conditions can also exhibit dynamic behavior via local structural fluctuations or by transiently sampling alternative (higher energy or activated) conformations. In many cases, such activated (non-native) states are functionally important despite their low Boltzmann weight. Realization of the importance of transient non-native protein structures for their function has not only greatly advanced our understanding of processes as diverse as recognition, signaling, and transport, but has also had profound practical implications, particularly for the design of drugs targeting specific proteins. Because of their transient nature, these non-native protein conformational states present a great challenge vis-à-vis detection and characterization. This chapter presents a concise introduction to an array of techniques that are used to study the structure and behavior of these "equilibrium intermediate states". We begin our discussion by considering protein ion charge-state distributions in electrospray ionization mass spectra as indicators of protein unfolding. We then proceed to various trapping techniques that exploit protein reactivity to reveal structural details of various non-native states. We conclude this chapter with a detailed discussion of hydrogen exchange, arguably one of the most widely used methods to probe the structure and dynamics of non-native (partially unstructured) protein states.

5.1. DIRECT METHODS OF MONITORING EQUILIBRIUM INTERMEDIATES: PROTEIN ION CHARGE–STATE DISTRIBUTIONS IN ESI MS

5.1.1. Protein Conformation as a Determinant of the Extent of Multiple Charging in ESI MS

As discussed in Chapter 3, the idea of using the ESI process as a means to generate (bio)polymer ions in a form suitable for MS analysis, predates the triumphant entry of this ionization technique into mainstream biological MS by some time. One particular feature of ESI that seemed to be a major

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

impediment for its application to macromolecules was multiple charging, a phenomenon that seemingly complicated the appearance of mass spectra by crowding them with multiple ion peaks corresponding to the same analyte. In retrospect, it seems almost ironic that the multiple charging of macromolecules in ESI MS is now viewed not as a liability, but as an extremely valuable asset, which provides an important new dimension to the analysis of biopolymers. In fact, it was not long after the acceptance of ESI MS as a tool for measuring masses of macromolecular ions, that observations were made linking the dramatic changes of ionic charge-state distributions to protein denaturation in solution. Chait and co-workers (1), and later Loo et al. (2), observed dramatic changes in the charge-state distributions of protein ions resulting from the changes in solvent composition that were known to induce protein denaturation.

Natively folded proteins, by definition, have stable compact cores sequestered from the solvent, and undergo ESI to produce ions carrying a relatively small number of charges. This occurs because the compact shape of a tightly folded polypeptide chain in solution does not allow the accommodation of a significant number of protons on its surface upon transition from solution to the gas phase. For this reason, ion peaks in the ESI mass spectra of proteins in aqueous solutions at a neutral pH typically dominate the high m/z regions of the spectra and are almost always characterized by having a narrow distribution of charge states (Fig. 5.1*a*). Unlike folded proteins, conformers that are either partially or fully unfolded in solution give rise to ions carrying a larger number of charges and their charge-state distributions are significantly broader (e.g., ion peaks labeled with open circles in Figure 5.1*b* and *c*). This occurs because once a protein loses its compactness upon denaturation (or unfolding), a significantly larger number of charges can be accommodated on its surface. Native and non-native protein states often coexist at equilibrium under mildly denaturing conditions. In such situations, protein ion charge-state distributions become bimodal, reflecting the presence of both native and denatured states. Dramatic changes of protein charge-state distributions therefore often serve as gauges of large-scale conformational changes (3).

Evolution of the protein ion charge-state distributions, shown in Figure 5.1, reflects conformational changes caused by the protein environment. However, there are many other situations when charge-state distributions reveal conformational transitions. Soon after the initial reports by Chait and co-workers (1) and Loo et al. (2), Fenselau and coworkers (4) demonstrated that protein ion charge-state distributions also provide information on protein conformational changes in solution induced by cofactors (e.g., metal ions) under near-native conditions. In these experiments, addition of divalent metals (Zn^{2+} or Cd^{2+}) to a solution of a small protein, apo-metallothionein, not only resulted in increases of the measured masses of the protein ions

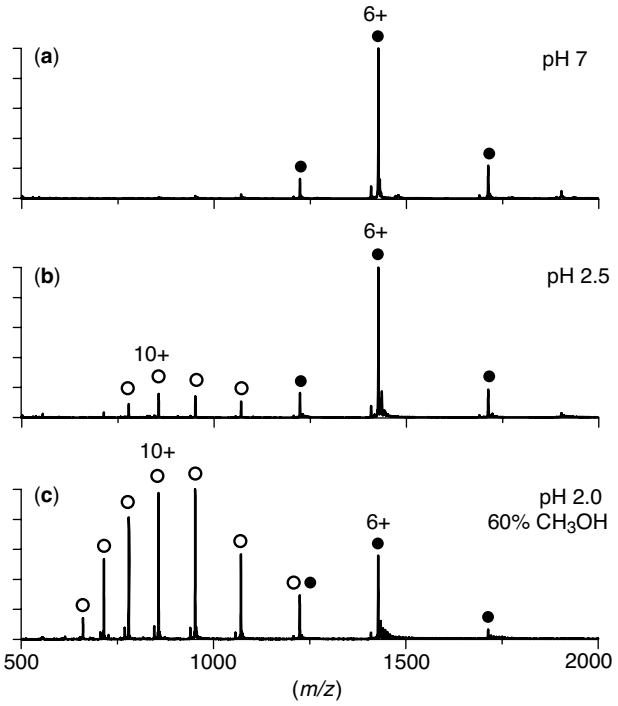


Figure 5.1. The ESI mass spectra of ubiquitin acquired under near-native (*a*, 10-mM $CH_3CO_2NH_4$, pH adjusted to 7) and denaturing conditions (*b*, 10-mM $CH_3CO_2NH_4$, pH adjusted to 2.5 with CH_3CO_2H; (*c*), low ionic strength, pH 2.0, 60% CH_3OH (methanol). Protein ions corresponding to the "native" and "non-native" (denatured) states of the protein are labeled with shaded and open circles, respectively.

[consistent with binding of a requisite number of (up to seven) metal cations to the protein], but also induced noticeable shifts in their charge-state distributions. The latter were interpreted as a manifestation of formation of a compact protein structure cemented by coordination of the metal ions by distant cysteine residues (4).

5.1.2. Detection and Characterization of Large-Scale Conformational Transitions by Monitoring Protein Ion Charge-State Distributions in ESI MS

The ability of these relatively simple measurements to probe conformational changes in proteins has been utilized extensively in the past two decades to study protein unfolding and conformational heterogeneity, as well as monitoring large-scale conformational changes within proteins in solution. Protein ion charge-state distributions provide an effective means of monitoring changes in protein conformation induced by changing solvent pH and composition (5–10), as well as solution temperature (11). This approach can also be used to study the effect of metal ion (12–19) and small organic ligand (20) binding on protein conformational stability. Likewise, influence of conformational heterogeneity on protein–protein (21,22) and protein–DNA (23) binding

can similarly be evaluated by monitoring protein ion charge-state distributions. Analysis of protein ion charge state distributions in ESI MS has also been used to determine the influence of mutations (24) and chemical modifications (25,26) on protein stability. Changes in the protein ion charge-state distributions can also be used to detect the "loosening" of the protein tertiary structure in solution following reduction of disulfide bonds (27).

This tool is now being actively evaluated in the biopharmaceutical sector, where it has been shown to be able to discriminate among monoclonal antibodies produced by different cell lines (28), as well as to detect loss of conformational stability in protein drugs triggered by non-enzymatic post-translational modifications (29,30). An example of such work is shown in Figure 5.2, where alkylation of a single free cysteine residue in interferon-β1a (IFN) leads to the appearance of high charge-density protein ions in ESI mass spectrum acquired under near-native conditions, while only low charge-density species could be observed for the intact protein. This example highlights the unique advantage of protein conformation analysis by ESI MS, which produces two independent readouts in a single measurement: mass, which can be used to make a distinction among various species present in solution, and ionic charge distributions, which are used to assess the integrity of protein conformation. This also

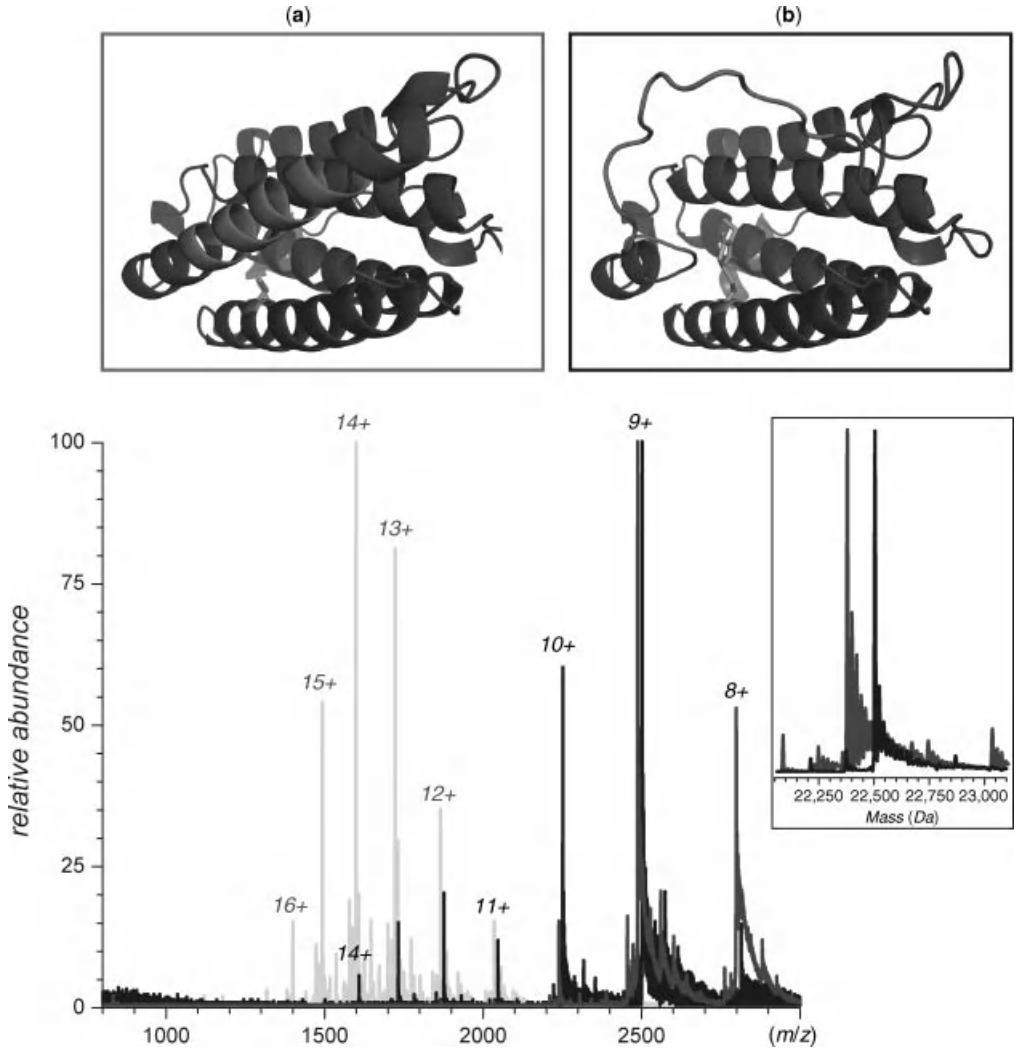


Figure 5.2. The ESI MS of intact (gray trace) and NEM alkylated interferon-β1a (black) in 100 mM ammonium acetate. The inset shows limited heterogeneity of interferon-β1a due to the presence of several glycoforms. The light gray trace shows an ESI mass spectrum of intact IFN acquired under strongly denaturing conditions. The two structures show (*a*) the native conformation of the protein and (*b*) the proposed destabilized form where covalent modification of a free Cys residue (shown as a stick model) results in melting of a helix. [Adopted with permission from (32). Copyright © 2010 American Society for Mass Spectrometry].

allows the protein ion charge-state distribution analysis to be used in complex multicomponent systems (21,22,31).

While conformational integrity is clearly compromised in the modified form of interferon-β1a, shown in Figure 5.2, a significant proportion of protein molecules appear to maintain compact conformation (represented by charge states 8+ through 10+). The rest of the modified protein molecules are less compact (11+ through 14+), but even they retain at least some residual structure. Indeed, acid unfolding of the intact protein leads to the appearance of ionic species in ESI MS, whose charge density is noticeably higher than that of the modified protein ions. Charge-state distributions representing the unfolded protein (light gray trace in Fig. 5.2) and the partially structured conformation (black trace, charge states 11+ through 14+) have a significant overlap. In this particular case, a distinction between the two states can be easily made based on the mass difference between the two forms of the protein. However, if the two non-native forms of the same protein were present in solution simultaneously, it would be very difficult to tease apart contributions made by each species to the total ionic signal.

Interferon, of course, is not the only protein having partially structured non-native states. In fact, most proteins have multiple non-native states, not just folded and unfolded conformations. Although in some cases the existence of at least one *intermediate state* (here we will use this term as a general descriptor of any non-native partially structured protein conformation) can be inferred from the charge-state distributions (33), it is possible to have multiple intermediate states coexisting in solution under equilibrium, some of which may have only minor structural differences. In most cases, such subtle conformational changes do not lead to significant variations in the overall shape of the protein and, therefore, in the average number of charges accommodated by the protein ion upon its desorption from solution. As a result, the charge-state distributions corresponding to various protein conformational isomers may be unresolved or poorly resolved (i.e., two or more different conformers may give rise to ions carrying the same number of charges). An important exception is the native conformation, whose ionic signal in most cases can be distinguished with relative ease from that of non-native states. Therefore, changes in the protein ion charge-state distributions are frequently regarded as qualitative indicators of re- or denaturation that cannot provide much information beyond gain or loss of the native fold.

This problem can be addressed, at least in some cases, using a procedure that utilizes chemometric tools to extract semiquantitative data on multiple protein conformational isomers coexisting in solution under equilibrium (34,35). Experiments are carried out by acquiring an array of spectra over a range of both near-native and denaturing conditions to ensure adequate sampling of various protein states and significant variation of their populations within the range of experimental conditions. The total number of protein conformers sampled in the course of the experiment can be determined by subjecting the set of collected spectra to singular value decomposition, SVD (36). The ionic contributions of each conformer to the total signal can then be determined by using a supervised minimization routine (Fig 5.3). Application of this method to several small model proteins has yielded a picture of protein behavior consistent with that based on the results of earlier studies that utilized a variety of biophysical techniques (35). More recently, this technique was used successfully to uncover copopulated non-native states of β_2-microglobulin (8), as well as conformational heterogeneity of several intrinsically disordered proteins (17,37-39).

One needs to be aware, however, that protein shape (compactness) is not the only factor that determines the appearance of charge-state distributions in the ESI mass spectra. Any factor that affects either the desorption process or the gas-phase ion chemistry may have a profound effect on the appearance of the spectra. Solvent composition is perhaps the most important parameter that can affect protein charge-state distributions, by either altering the conditions of ion desorption (by altering solvent dielectric constant, surface tension, etc.) (40,41) or by supplying proton transfer reagents to the gas phase (40,42). Changes in solvent composition may also indirectly affect charge-state distributions, for example, via formation of protein–anion adducts in solution that dissociate in the gas phase without charge partitioning, resulting in apparent charge-state reduction (43). Care must be exercised, therefore, in order to avoid overinterpretation of the experimental data based on measurements of protein ion charge state distributions. In many cases, it is possible to judiciously select a range of solution conditions that introduce no variation in the protein ion charge-state distributions other than those related to shifts of equilibria among various protein conformers.

5.1.3. Detection of Small-Scale Conformational Transitions by Monitoring Protein Ion Charge-State Distributions

Unlike large-scale unfolding, small-scale conformational transitions are much more difficult to detect, as they result in much more subtle changes of protein ion charge-state distributions. The resulting changes of the extent of multiple charging in this case could be smaller than the charge-state shifts caused by other external factors discussed in Section 5.1.2. Nevertheless, it is possible to choose experimental conditions in such a way that the influence of gas-phase ion chemistry on protein ion charge-state distributions would remain near constant over the entire range of conditions used in ESI MS experiments. For example, if the objective of a study is to assess conformational

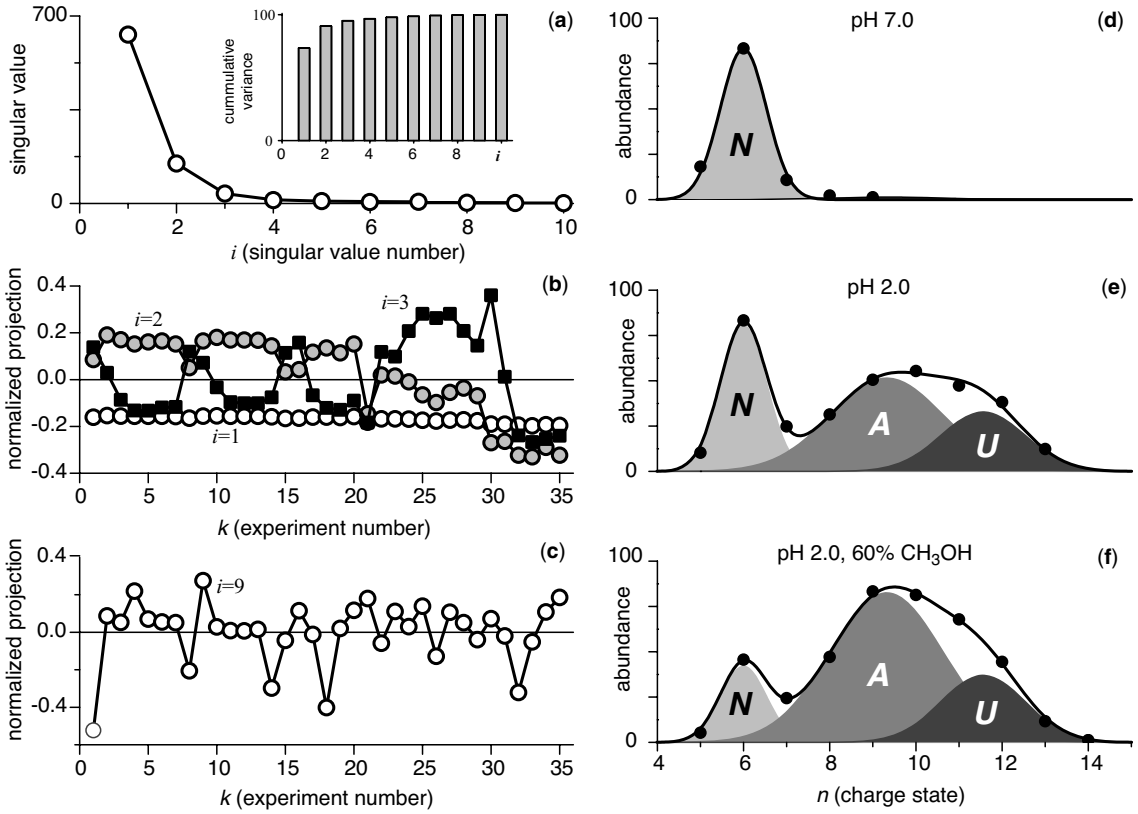


Figure 5.3. Factor analysis of an array of ESI mass spectra recorded at various stages of acid- and alcohol-induced unfolding of ubiquitin (representative raw data are shown in Fig. 5.1). The total number of protein states is determined by a singular value decomposition of the entire data matrix of dimension $N \times K$ (charge state $\times$ experiment number). While SVD produces a total of 10 singular values, only the first three are required to account for 95% of variation exhibited in the spectra within the entire range of conditions (*a*). "Abstract" solutions corresponding to these three "significant" singular values show consistent behavior throughout the entire range of conditions (*b*), while the "insignificant" singular values randomly oscillate around zero (an example is shown on panel *c*). Contributions of the three protein states (assigned as ***N***ative, ***A***-state and fully ***U***nfolded) to the total ion current are determined using supervised optimization (three examples are shown on *d*-*f*). [Adapted with permission from (35). Copyright © 2003 American Chemical Society.]

heterogeneity of a protein over a wide pH range, electrolyte selection for the protein solution becomes critically important. In a common situation, when aqueous A^-BH^+ solution is acidified by incremental additions of AH, the pH should not be lowered below the pK_a of AH (43). This places nearly inhibitive restrictions on the studies of acid-induced unfolding when employing two of the most popular "native" ESI MS solvent systems, namely, ammonium bicarbonate (p$K_{a1} = 6.4$) and ammonium acetate (p$K_a =$ 4.7). To expand the pH range over which studies of conformational dynamics can be carried out, the traditional acetate- or bicarbonate-based solvent systems can be replaced with those that have a very weak basic component A^- (i.e., whose conjugate acid AH is rather strong). For example, replacing ammonium acetate with ammonium formate lowers the pH at which the apparent charge reduction occurs by one pH unit (pK_a of HCO_2H is 3.7). The pH limit in the studies of acid-induced conformational changes could be pushed much lower, down to pH 1, should the ESI MS measurements be carried out using ammonium trifluoroacetate ($CF_3CO_2NH_4$) solutions, whose pH is adjusted with an acid.

We have used this approach to study the pH dependent inactivation of pepsin (44), an acidic protease, which starts to lose its enzymatic activity at pH > 2, in a process that involves a relatively small-scale conformational transition (45,46). Although two conformational transitions at pH 2.5 and 7.5 are readily revealed in fluorescence experiments (Fig. 5.4*b*), high charge-density ions of pepsin appear in ESI MS only at pH 7.5 (Fig. 5.4*a*). This suggests that the first transition is not accompanied by global unfolding of the protein; indeed, other spectroscopic measurements suggest the protein remains compact at pH < 7 (44). Despite the absence of high-charge-density pepsin ions in this pH range,

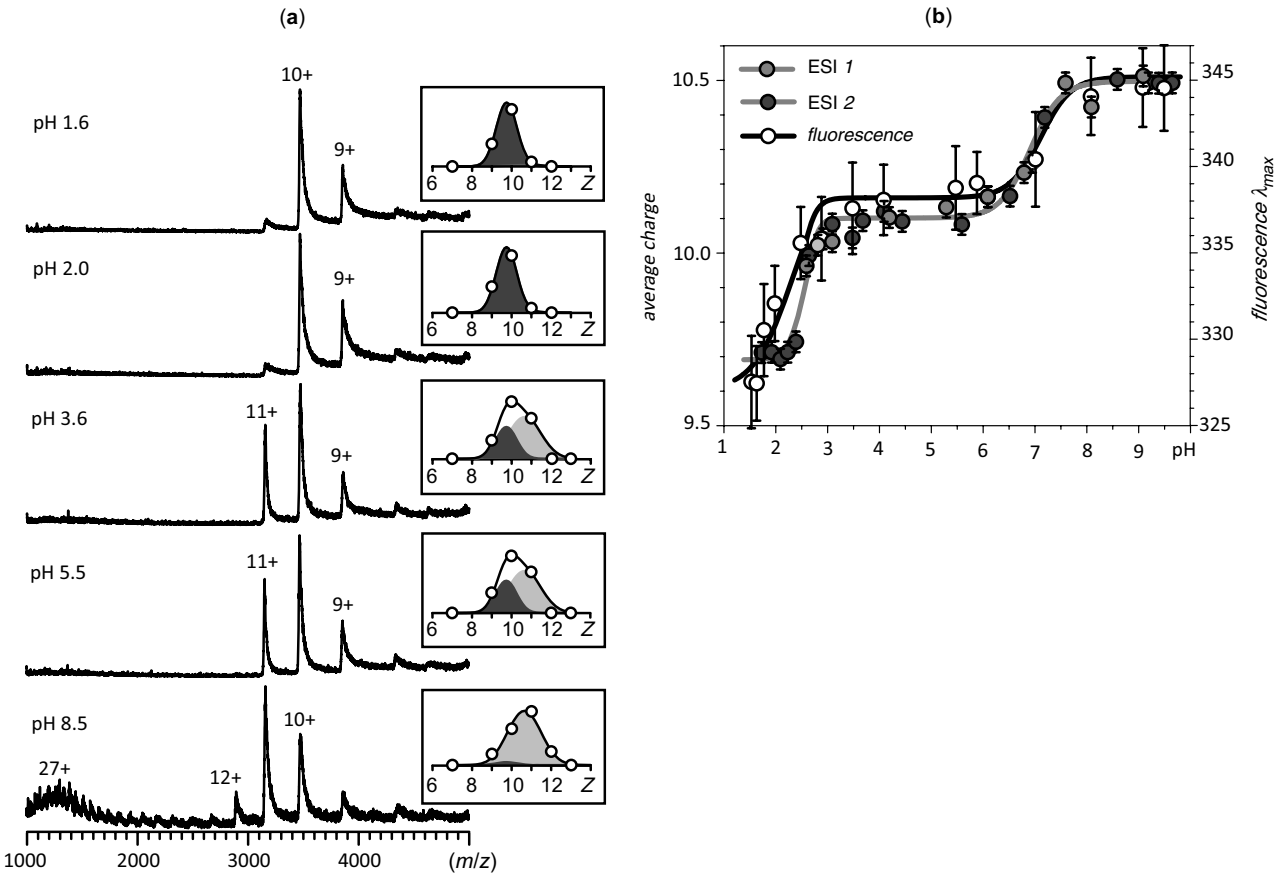


Figure 5.4. (*a*) Representative ESI mass spectra of pepsin acquired in 10-mM $CF_3CO_2NH_4$ solutions whose pH levels were adjusted as indicated on the panels using HCO_2H or NH_4OH. The boxed insets show the results of fitting the intensity distributions of low charge density pepsin ions with two basis functions (Gaussian curves), representing active (dark gray) and compact inactive (light gray) conformations of the protein. (*b*) Evolution of the maximum fluorescence wavelength of pepsin in solution and the average charge of low charge density pepsin ions in ESI MS as a function of pH (the two ESI data sets were acquired on different days). [Reproduced with permission from (44). Copyright © 2007 American Chemical Society].

the charge-state distribution clearly evolves, suggesting the presence of two compact species in solution. Careful analysis of pepsin ion charge-state distribution reveals the presence of two transitions, in excellent agreement with the fluorescence data (e.g., compare the average ionic charge and λ_{max} curves in Fig. 5.4*b*).

When executed with due care, ESI MS experiments provide a means to observe small-scale conformational transitions that do not result in significant loss of higher order structure. In Chapter 9, we will consider another example of using ESI MS to detect small-scale conformational changes that trigger protein aggregation. An intriguing question that might be asked here relates to the possibility of extracting even more information from the analysis of protein ion charge-state distributions. In Chapter 4, we already discussed the possibility of obtaining estimates of solvent-exposed surface areas of natively folded proteins based on the extent of their multiple charging in ESI MS. Since deconvolution of charge-state distributions discussed in this and the preceding sections also provides an opportunity to determine average ionic charges corresponding to non-native conformers, could it be used to determine their physical dimensions in solution in a fashion similar to that used for natively folded proteins (47)? A recent analysis of literature data by Grandori and co-workers (48) suggests that a correlation exists between the extent of multiple charging of fully unstructured proteins and their calculated surface areas in solution (Fig. 5.5), and the form of this relationship is very similar to (but not exactly the same as) that for natively folded proteins. Clearly, more work is needed in order to establish with certainty the existence of a universal charge/surface relationship. If such a relationship does indeed exist, analysis of protein ion charge-state distributions will provide new and important dimension in characterizing non-native protein states by allowing estimates of their physical dimensions to be deduced directly from the ESI MS measurements.

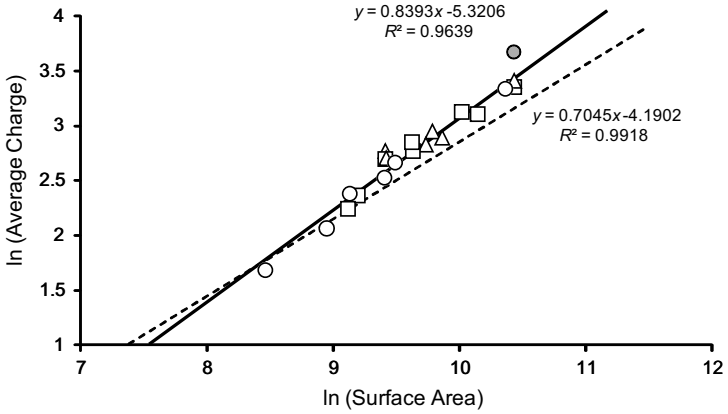


Figure 5.5. Charge-to-surface correlation for denatured globular proteins and intrinsically disordered proteins (IDPs) in ESI MS (literature compilation). The logarithm of the experimental average charge in the positive-ion mode is plotted against the logarithm of the solvent-accessible surface area calculated by ProtSA (circles, IDPs under denaturing conditions; triangles, globular proteins under denaturing conditions; squares, IDPs under nondenaturing conditions) and the best linear fit (solid line). The dashed line shows the result of a similar analysis for globular proteins under nondenaturing conditions (data points not shown). [Adapted with permission from (48). Copyright © 2011 American Chemical Society.]

5.1.4. Pitfalls and Limitations of Protein Ion Charge-State Distribution Analysis

This section would be incomplete without considering the limitations of charge-state distribution analysis, as well as common pitfalls that are associated with experimental work using this technique. We already mentioned the potential influence of gas-phase processes on the ionic charges. Another interesting phenomenon that may complicate the analysis of protein conformational heterogeneity based on the appearance of ionic charge-state distributions in ESI MS is the artificial enhancement of the ionic signal of (partially) unfolded protein species in ESI MS, which may give rise to false-positive signals of protein unfolding (49). This phenomenon is observed when significant quantities of protein complexes or aggregates exist in solution. As already discussed in Chapter 4, dissociation of such complexes in the gas phase usually proceeds via asymmetric charge partitioning, where a single polypeptide chain is ejected from a metastable complex and carries a disproportionately high number of charges (50). An example of such a process was shown in Figure 4.4, where mild collisional activation of the 14-meric protein complex GroEL in the ESI interface resulted in ejection of a highly charged monomeric ion from the complex. The monomeric products of dissociation populate the low end of the *m/z* scale (average charge 29+), and their presence in the mass spectrum alongside the low-charge-density monomeric ions (average charge 16+) falsely suggests the existence of unstructured monomers in solution at equilibrium with the natively folded species.

Above and beyond ion molecule charge transfer and ion dissociation reactions, even selection of data acquisition parameters in ESI MS may (and almost always does) introduce bias in the ionic charge-state distributions by altering the transmission/detection efficiency as a function of *m/z*. Obviously, analysis of charge-state distributions of protein ions in ESI MS provides reliable information on protein conformational dynamics only if the observed changes in the extent of multiple charging are due to changes of protein compactness in solution. Protein ion charge-state distributions are also very sensitive to the way that the ions are produced (51), hence the frequently observed variations, not only across various instrumental platforms, but even with a single mass spectrometer based on specific physical parameters employed to generate an ESI signal.

The unique ability of ESI MS to visualize individual conformers often leads to a temptation to determine the relative abundance of various protein states in solution based on ESI MS data, but are these fractional concentrations of various protein states in solution proportional to the signal intensity of ions representing these conformers in ESI MS? Many early studies implicitly assumed that the answer to this question is "yes", although some reservations were also expressed (34). A recent study by Kuprowski and Konermann (52) explored the relationship between the fractional concentrations of different conformers and their respective ionic signals. Convincing evidence was presented that non-native polypeptide chains generate higher signal response compared to the natively folded species. In some unfavorable cases, the ionic intensity ratios may deviate from the actual concentration ratios by as much as two orders of magnitude. The higher ionization efficiency of unfolded proteins is attributed to their increased hydrophobicity, which increases their surface activity. Consequently,

these species are much more likely to be located at the solvent–air interface in the charged droplets produced by electrospray and will have a much higher probability to be transferred to the ion-producing progeny droplets during Coulombic explosion events. Fortunately, the study also found that suppression of the ionic signal corresponding to the natively folded protein species can be minimized if the experiments are carried out in a charge-surplus regime, that is, at relatively low-protein concentrations (52).

Finally, it should be noted that not all concerns that are frequently expressed over the validity of protein conformational analysis based on charge-state distributions are well founded; in fact, some of the perceived limitations of this technique are imaginary. For example, it was (and still is) not uncommon to hear an argument that the extent of multiple charging of unfolded polypeptides in ESI MS must be limited by the number of available ionizable sites. As discussed in Chapter 3, multiple charging of proteins in ESI MS most commonly occurs in the positive-ion mode in the form of protonation (the attachment of multiple protons, H^+), although other types of polycation formation can also occur. Formation of polyanionic species in negative ion ESI MS usually proceeds via deprotonation of polypeptides. It is the intimate involvement of protons in generating multiply charged ions of biopolymers that led to suggestions that the charge-state distributions observed in ESI MS must reflect the cumulative charge on the acidic and basic residues in solution (53,54). In this view, the large-scale conformational dynamics of the polypeptide chains in solution influenced the charge-state distributions indirectly, by modulating the pK_a values of individual amino acid residues.

Although the acid–base paradigm is a convenient way of thinking of protein ions in solution, it is a very poor predictor of the extent of multiple charging of proteins in ESI MS. Fenselau and co-workers (55) demonstrated that both polycationic and polyanionic species could be generated from a protein solution at any given pH, proving that evolution of the charge distribution reflects conformational transitions, rather than titration of basic or acidic side chains. The notion of solution-phase acid–base chemistry as a determinant of the extent of multiple charging in ESI MS was abandoned a long time ago; however, acid–base chemistry is still frequently invoked as a factor influencing protein ion charge-state distributions. For example, it is often argued that the extent of multiple protonation of an unstructured polypeptide chain should be limited by the number of basic residues in its sequence (56). Specifically, proton affinity of the solvent molecules is suggested to provide a cut-off level for amino acid residues that can be protonated in the gas phase (42). Pepsin, mentioned in Section 5.1.3, is an ideal system to test these assumptions. It contains 41 acidic, but only four basic residues, so it would be tempting to argue that there are simply not enough basic sites in this protein to afford adequate protonation of the unfolded polypeptide chain in the gas phase. Despite the small number of basic sites, it is evident from Figure 5.4 that pepsin ions accumulate up to 11 positive charges when desorbed from aqueous solutions at near-native pH 2 (the native environment of pepsin, gastric juice, is extremely acidic).

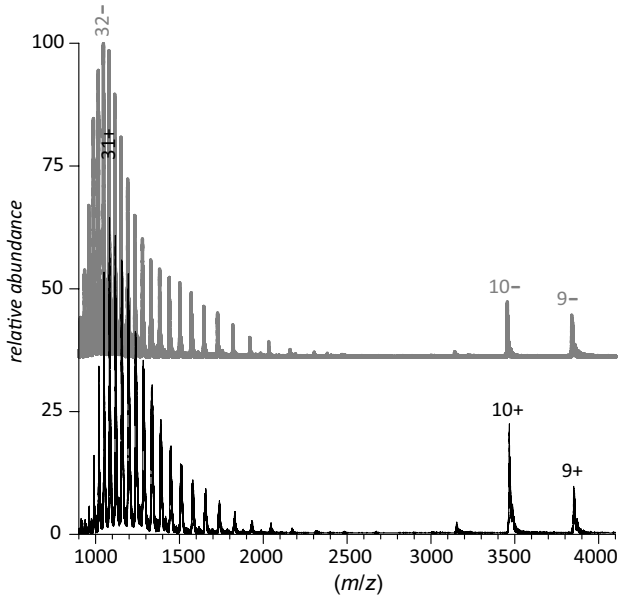


Figure 5.6. The ESI mass spectra of porcine pepsin acquired under denaturing conditions (20 mM ammonium acetate, pH adjusted to 9.5, 50% CH_3OH by volume) acquired in the positive (black trace) and negative (gray trace) ion modes. [Adapted with permission from (33). Copyright © 2008 American Society for Mass Spectrometry.]

Large-scale unfolding of pepsin occurs in neutral and basic solutions. Therefore, it is not surprising to observe a bimodal charge-state distribution of pepsin ions in ESI MS acquired under these conditions (Fig. 5.6). Importantly, the most abundant ion peak in the mass spectrum corresponds to a protein molecule accommodating 31 protons, more than six times the number of basic sites on the polypeptide chain! It is quite remarkable that a polypeptide chain, which contains only four basic residues, can accommodate so many protons. It is perhaps even more surprising that the extent of multiple charging of pepsin polyanions (in negative ion ESI MS) is very similar to that of multiply protonated species in positive ion ESI MS (cf. the two traces in Fig. 5.6), despite more than a 10-fold excess of acidic residues in this protein. The fact that the polycationic charge-state distribution of pepsin closely mirrors that of the polyanionic species (despite the abundance of acidic residues and the extreme deficiency of basic ones) strongly suggests that the extent of multiple charging is indeed determined by the protein geometry in solution, not the number of available basic sites.

5.2. CHEMICAL LABELING AND TRAPPING EQUILIBRIUM STATES IN UNFOLDING EXPERIMENTS

5.2.1. Characterization of Solvent-Exposed Surfaces with Chemical Labeling

Although chemical cross-linking and selective chemical modifications are typically used to assess native protein structures, the same approaches can be used in principle to characterize nonhomogeneous protein structures (e.g., equilibrium intermediate states). There are, however, very few examples of using either chemical cross-linking or selective labeling in conjunction with MS to characterize protein conformational heterogeneity under equilibrium. One possible reason is that the structural information obtained in such experiments is ensemble-averaged, which often makes data interpretation rather difficult. This was exemplified by Craig et al., (57) who used a photochemical reagent to characterize the degree of unfolding of a small protein α-lactalbumin. The urea-unfolded state of the protein was labeled 25–30% more than the native state due to an apparent increase of the solvent-accessible surface area. However, such an increase is still below the increment expected from the theoretical estimates of the solvent-accessible surface area of the random coil (100% was predicted based on theoretical calculations). This discrepancy was attributed to "residual structure" in the unfolded state. Another possible explanation would be presence of a partially structured intermediate state (e.g., a molten globular state) alongside the random coil, with the experiments providing the solvent-accessibility patterns averaged across the entire protein population.

More recently equally suspicious results were obtained with this technique, when diazirine (DZN) was used as a photochemical probe of structure of the acid-induced molten globular state of α-lactalbumin (58). Although the extent of labeling in this protein under acidic conditions was higher than that under native conditions (consistent with loss of structure), it also exceeded the level of $:CH_2$ tagging observed under conditions favoring complete loss of structure. Although the extent of the labeling increase was uneven across various protein segments, all of them showed an increase in reactivity (Fig. 5.7). This observation seems counterintuitive, as residual structure in the molten globular

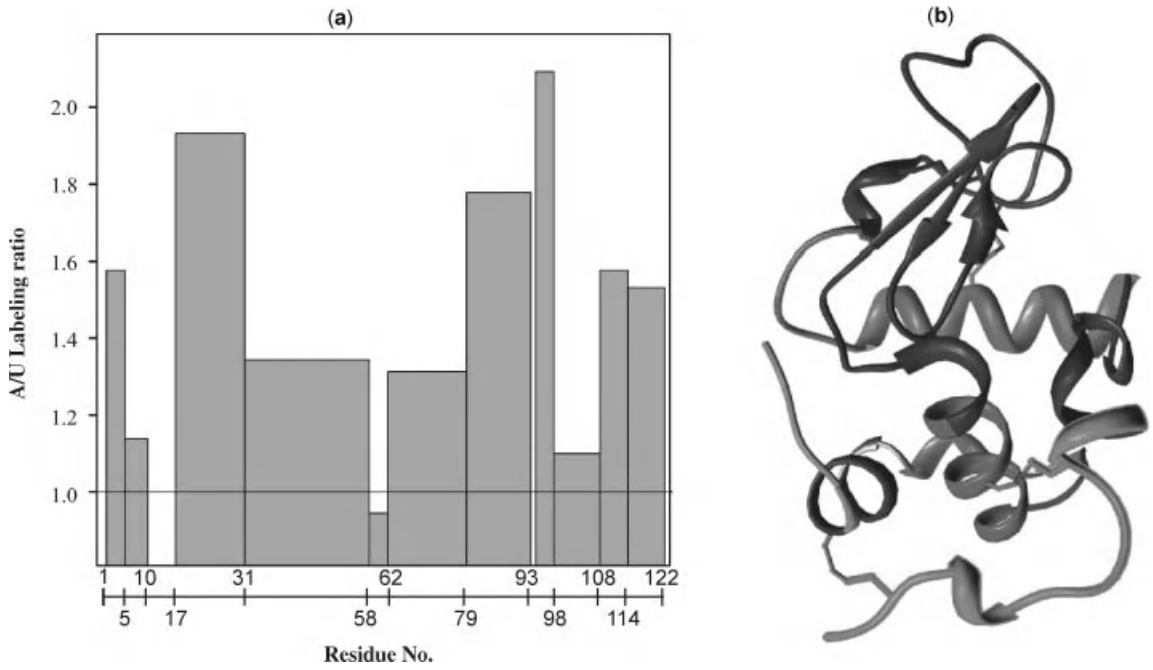


Figure 5.7. (*a*) Regiospecific $:CH_2$ labeling pattern of the acid-stabilized molten globule (A-state) of bovine α-lactalbumin relative to that of the unfolded protein (U-state). The A/U labeling ratio for each tryptic peptide was measured as the $:CH_2$ labeling yield of a sample modified in the A-state relative to the corresponding value for a sample of equivalent mass modified in the U-state. (*b*) α-Lactalbumin backbone colored according to the $:CH_2$ labeling change in the A-state relative to the U-state. Tryptic peptides showing the highest labeling enhancement (A/U values in the range 1.8–2.1) are shown in red (residues 17–31, 80–93, and 95–98); those showing intermediate values (in the range 1.5–1.6) are shown in cyan (residues 1 - 5 + 109–114 and 115–122); those presenting minimal enhancements (A/U values <1.4) are shown in blue (residues 6–10, 32–58, 59–62, 63–79, and 99–108). The disulfide bridges are shown in green. [Reproduced with permission from (58).] (See color version of the figure in Color Plate section)

state would be expected to decrease the reactivity of the polypeptide chain toward $:CH_2$, opposite to what was observed. This phenomenon was rationalized in terms of the ability of the hydrophobic core of the molten globular state to effectively partition the nonpolar DZN, thereby increasing its local concentration (58). This example clearly shows that the structure–reactivity relationship is far from straightforward, and care needs to be exercised when designing such experiments and interpreting their results.

An additional disadvantage of chemical labeling as a technique to study dynamic protein structures is that slow diffusion of a bulky labeling reagent to its target site may prevent efficient labeling. Even if the kinetics of the labeling reaction itself are very fast, the efficiency of a trapping reaction will be significantly limited by the slow diffusion of reactant (covalent modifier) through the protein solution to the transiently exposed reactive site on the protein. However, there are several notable exceptions to this rule, where the diffusion limit can be defeated and chemical modifications can be used to study conformational dynamics. The first two employ an omnipresent modifier (solvent), which either reacts selectively with non-native protein states, but is inert toward the native state, or else is totally unreactive in its ground state, but can be activated on a very short time scale. The former is, of course, the well-known and extremely popular technique of hydrogen–deuterium exchange (HDX), which will be discussed in Section 5.3. The latter is fast nonspecific labeling, whose applications to characterization of static protein structures have been discussed already in Chapter 4. While laser-induced water photolysis appears to be the most popular way to generate $HO^{\bullet}$ radicals within a very narrow time window suitable for probing the structure of transiently populated intermediate states (59,60), time-resolved radiolytic footprinting can also be used for this purpose [reviewed in (61)]. These methods, as well as a recently introduced γ-ray mediated oxidative labeling (62) can also be used to characterize equilibrium intermediate states, as less structured conformers exhibit higher labeling levels than tightly folded species (62–64). However, it should be noted that the outcome of such measurements frequently depends not only on the protein conformation, but also on the solvent system used in the experiments. Indeed, many popular chaotropic agents are efficient radical scavengers, and therefore modulate the extent of protein oxidative labeling (62). Applications of these techniques to investigate kinetics of protein folding and assembly are discussed in Chapter 6.

5.2.2. Exploiting Intrinsic Protein Reactivity: Disulfide Scrambling and Protein Misfolding

Another approach that can be used to probe the structure of non-native states without the need for external covalent modifiers relies on the internal chemically active groups within the protein. As already discussed in Chapter 4, cysteine residues have an intrinsic propensity to form stable bonds in the native conformation (65,66), and monitoring disulfide-bond formation reactions can also be used to study protein conformation and dynamics. The reactions include oxidation, reduction, and reshuffling, which are all based on thiol—disulfide exchange, in which the thiolate anion R_1S^- displaces one sulfur of the disulfide bond R_2SSR_3 (65). Thiol–disulfide exchange reactions are attenuated by a variety of factors, including local electrostatic interactions and structural propensities. Stable tertiary structure is another important determinant, as it locks in the native disulfide bonds. Thus, regeneration of the disulfide bond pattern is often used to monitor the progress of protein folding (67). Traditional techniques of determining the disulfide bond pattern rely on proteases to cleave the polypeptide backbone between the cysteine residues. In some cases, the disulfide bond pattern within trapped folding intermediates can be determined using a combination of proteolysis and MS analysis (68). However, application of this procedure to proteins containing a large number of cysteine residues can be a very challenging task, particularly when the cysteine residues are located close to each other in the sequence.

We discussed some approaches to mapping disulfide bonds in natively folded proteins in Chapter 4; many of these same procedures can be used to obtain non-native disulfide patterns. In addition, the disulfide bond pattern in partially reduced proteins can be obtained using cyanylation, which induces specific backbone cleavages at the N-terminal side of modified (*S*-cyano-cysteine) residues (69–71). The backbone cleavage step is followed by complete disulfide reduction of the peptide mixture. The reduced peptides are separated by high-performance liquid chromatography (HPLC) and analyzed by matrix-assisted laser desorption–ionization (MALDI) or ESI MS. This procedure can be applied readily to study late protein-folding intermediates by quenching folding and trapping the disulfide intermediates under conditions that minimize disulfide scrambling (i.e., sulfhydryl–disulfide exchange) in solution (72,73). An alternative procedure for mapping proximal cysteine residues in partially unstructured proteins utilizes bis-thiol reagents (derivatives of arsenous acid, e.g., melarsen oxide and pyridinyl-3-arsonous acid) to link neighboring cysteine residues in reduced proteins (74,75). Bis-thiol selective derivatization of intermediate states is principally different from mono-thiol trapping strategies, since bis-thiol modifications actually cross-link closely spaced cysteine residues pairwise. As a result, chemical reduction does not result in protein unfolding. Two additional advantages of this experimental strategy are (1) relatively large protein mass increase upon single modification and (2) absence of any side reactions even with a high molar excess of the reagent. As is the case with other chemical-labeling techniques, the identity of the modified residues can be

established using a combination of proteolysis, separation and MS analysis. Among the more recent developments in this field, we note a successful application of electron-based ion-dissociation techniques to obtain non-native disulfide patterns in large proteins, such as monoclonal antibodies and antibody-based fusion proteins (76).

5.3. STRUCTURE AND DYNAMICS OF INTERMEDIATE EQUILIBRIUM STATES BY HYDROGEN EXCHANGE

5.3.1. Protein Dynamics and Hydrogen Exchange

Perhaps the major disadvantage of the experimental methods presented thus far is their inability to detect small-scale and/or rare dynamic events in solution. Here we use the term "small-scale dynamic events" as a general descriptor of local protein motions that alter structure of a small portion of the protein without affecting the rest. An example of such an event would be a local conformational fluctuation. As discussed in Chapter 1, even under native or near-native conditions, proteins sample non-native states via either transient fluctuations or even large-scale unfolding events. Such large-scale motions (partial or complete loss of the native protein structure) are very rare events, unless, of course, the protein under consideration is an intrinsically unstructured one. The small-scale events do not alter the overall structure enough to induce changes in the charge-state distributions, while their transient nature prevents efficient chemical trapping. Furthermore, if the Boltzmann weight of the putative non-native states is negligible, as is often the case under native conditions, they would not be detectable using either protein ion charge-state distributions or chemical-labeling techniques.

One of the potent experimental tools that has been widely used to study protein dynamics is hydrogen–deuterium exchange (HDX) (77). These measurements detect little or no contribution from the protein molecules in the ground state, providing an effective means to visualize transiently populated activated states (78,79). The pioneering work of Katta and Chait (80) first demonstrated the great potential of the HDX–ESI MS combination as a tool to probe conformational dynamics of small proteins. In the following years, the number and scope of applications of the HDX–ESI MS (and, more recently, HDX–MALDI MS) methodology to probe dynamics of biomolecules has expanded dramatically, catalyzed by spectacular technological improvements in soft ionization methods. ESI MS offers several important advantages over NMR, namely, faster time scale, tolerance to high-spin ligands and co-factors, ability to monitor the exchange in a conformer-specific fashion, as well as much more forgiving molecular weight limitations. The ability of ESI MS to handle larger proteins and their complexes is particularly important when compared to high-field NMR, which still has limited application for proteins larger than ~30 kDa. The practical upper mass limit of ESI MS, on the other hand, has yet to be established, as the bar is being continuously raised (81). Another significant advantage offered by ESI MS is its superior sensitivity, which allows many experiments to be carried out using only minute quantities of proteins. In many cases, this enables the studies of protein behavior at, or even below, endogenous levels. Importantly, because of the ability of ESI MS to desorb biomolecular species directly from aqueous solutions, as well as high data acquisition rates afforded by most mass analyzers, HDX–ESI MS measurements can often be carried out online, enabling in many cases real-time monitoring of protein dynamics.

In Chapter 4, we discussed briefly some basic aspects of HDX experiments relevant to probing "static" structures of proteins and protein assemblies. We noted that protein HDX involves two different types of reactions: (1) reversible protein unfolding that disrupts the hydrogen-bonding network, and (2) isotope exchange at individual unprotected amides. Since protein unfolding (either local or global) is a prerequisite for exchange at the sites that are protected in the native conformation, HDX reactions serve as a reliable and sensitive indicator of the unfolding events. "Protection" here means either involvement in the hydrogen-bonding network or sequestration from solvent in the core of the protein. However, it is important to remember that the organization and dynamic features of the hydrogen-bonding network are not the only determinants of the HDX kinetics. Even in the absence of any protection, the exchange kinetics of any labile hydrogen atom will strongly depend on the nature of the functional group. Thus, solvent-exposed η-hydrogen atoms of the Arg guanidine group would have the fastest exchange rates at neutral pH, exceeding those of backbone amide and Trp indole hydrogen atoms by more than an order of magnitude and almost three orders of magnitude, respectively. The pH dependencies of the cumulative intrinsic exchange rates for several types of labile hydrogen atoms, calculated based on the data compiled by Dempsey (82), are presented in Figure 5.8. The exchange rates are also influenced by sequence-neighboring residues via both inductive and steric blocking effects (84). Furthermore, since hydrogen exchange at unprotected sites is both acid and base catalyzed, the overall exchange rate constant will depend on solution pH. Exchange through catalysis by water has also been reported, and the "intrinsic exchange" rate constant is usually presented as (82):

$$k_{\text{int}} = k_{\text{acid}}[\text{H}^+] + k_{\text{base}}[\text{OH}^-] + k_{\text{W}} \qquad (5\text{-}3\text{-}1)$$

For most labile hydrogen atoms, base catalysis is more effective, exceeding the acid-catalyzed rate constant by four to eight orders of magnitude (82). Isotopic composition of

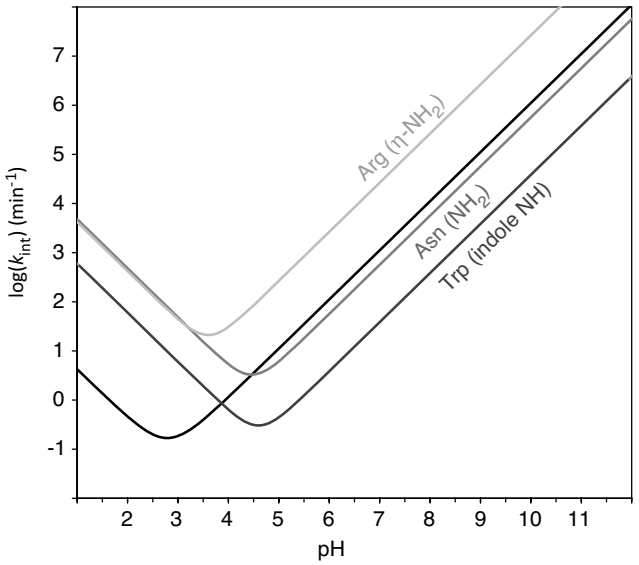


Figure 5.8. Intrinsic exchange rates of several types of labile hydrogen atoms as functions of solution pH calculated based on the data compiled in (83). Black line represents backbone amide hydrogen atoms.

both solvent and the protein also exerts a certain influence on the intrinsic exchange rates. Rate constants for the acid-catalyzed exchange of amide groups NH, ND (N^2H), and NT (N^3H) in ^1H_2O are essentially identical, but a solvent isotope effect doubles the acid-catalyzed rate in D_2O (83). Intrinsic exchange rate constants for base-catalyzed exchange in H_2O decrease slowly in the order NH > ND > NT, while the alkaline rate constant in D_2O appears to be very close to that in H_2O after making corrections for glass electrode pH artifacts and differences in water auto-protolysis constant (83). The presence of small exchange catalysts (e.g., phosphate, carbonate), as well small organic molecules with carboxyl and/or amino groups typically increases hydrogen-exchange rates from both the hydroxyl and amino groups of polypeptides (85). The presence of organic cosolvents also influences the intrinsic exchange rates, although this subject has not received much attention (82).

5.3.2. Global Exchange Kinetics in the Presence of Non-Native States: EX1, EX2, and EXX Exchange Regimes in a Simplified Two-State Model System

Backbone amide hydrogen atoms constitute a particularly interesting class of labile hydrogen atoms due to their uniform distribution throughout the protein sequence, which makes them very convenient reporters of protein dynamics at the amino acid residue level (proline is the only naturally occurring amino acid lacking an amide hydrogen atom). Therefore, it is not surprising that the majority of HDX experiments are concerned with the exchange of backbone amide hydrogen atoms. The mathematical formalism that is often used to describe HDX kinetics of backbone amides was introduced several decades ago and is based upon a simple two-state kinetic model (86):

$$\mathrm{ND(incompetent)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{ND(competent)} \xrightarrow{k_{int}} \mathrm{NH(competent)} \rightleftharpoons \mathrm{NH(incompetent)} \qquad (5\text{-}3\text{-}2)$$

where k_{op} and k_{cl} are the rate constants for the opening (unfolding) and closing (refolding) events that expose/protect a particular amide hydrogen to/from exchange with the solvent. Most of the data presented in this chapter will have a fully deuterated protein as a starting point of the exchange reaction, and HDX reactions are initiated by placing such a protein into a protiated buffer solution; therefore, we modified the original presentation of Eq. 5-3-2 by Hvidt and Nielsen (86), who used a fully protiated protein as a starting point. The intrinsic exchange rate constants of amide hydrogen atoms from the exchange-competent state k_{int} can be estimated using short unstructured peptides (87). The ND $\rightarrow$ NH transition is essentially irreversible, as HDX experiments are carried out in significant excess (typically 10–100-fold) of exchange buffer.

In most HDX studies, the exchange-incompetent state of the protein is considered to be its native state. The exchange-competent state is thought of as a non-native structure, which can be either fully unfolded (random coil) or partially unfolded (intermediate states). Alternatively, it can represent a structural fluctuation within the native conformation, which exposes an otherwise protected amide hydrogen to solvent transiently through *local unfolding* or *structural breathing* without unfolding (88,89). Transitions between different non-native states under equilibrium conditions are usually ignored in mathematical treatments of HDX. One of the reasons is that the majority of HDX measurements are carried out under native or near-native conditions. The Boltzmann weight of non-native states for most proteins under these conditions is very low and the transitions among such states do not make any detectable contribution to the overall HDX kinetics. Non-native HDX is now also experiencing a surge in popularity due to renewed interest in the structure and dynamics of intermediate protein states. An important advantage offered by ESI MS is its ability to directly visualize various protein states based on the difference in deuterium incorporation. As we will see below, a clear distinction between a native conformation and a (partially) unstructured state can be made under certain mildly denaturing conditions.

Unlike HDX NMR measurements, a typical HDX MS experiment provides information on *global protection patterns* by measuring the isotope content of the entire protein, rather than the exchange kinetics of individual amide hydrogen atoms. Nevertheless, interpretation of HDX MS data

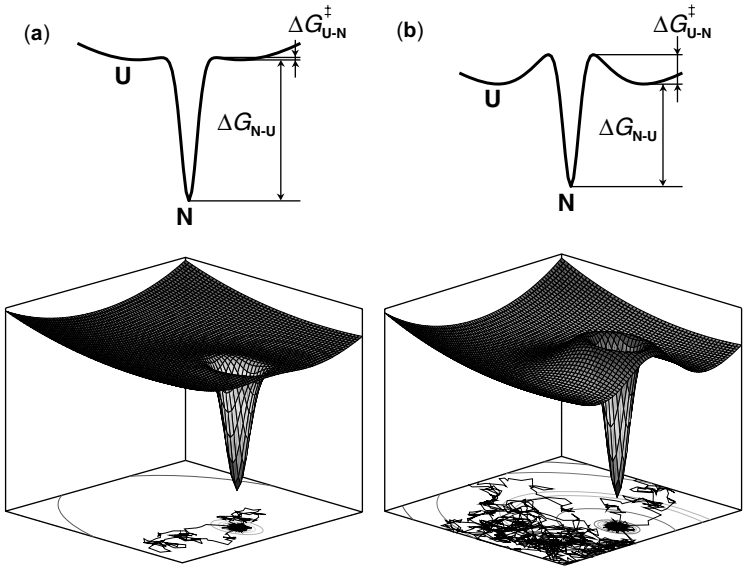


Figure 5.9. Minimalistic representations of a two-state protein system. The global energy minimum corresponds to an exchange-incompetent (native) state of the protein. The diffuse local minima (concentric to the potential well of the "native" state) represent an exchange-competent state (random coil). The excursions from the global minimum are rare due to the significant difference in energy (Boltzmann statistics) and very short-lived when the reverse activation energy barrier separating the local minima basin from the "native" potential is low (*a*). The protein molecules may become trapped in the "random coil" state for prolonged periods of time if the reverse activation energy barrier is significant (*b*).

often utilizes the kinetic model (Eq. 5-3-2) by making an implicit assumption that ND(incompetent) and ND(competent) represent *groups of amides*, rather than *individual amides* that become unprotected upon transition from one state to another. This simplistic view of protein behavior is illustrated in Figure 5.9, where the global energy minima (narrow deep potential wells) represent native protein conformations, while the diffuse shallow local minima represent the unstructured (exchange-competent) states. The potential wells representing global minima are narrow, allowing no conformational freedom for the protein (exchange-incompetent state). An exchange event can only occur if the protein molecule escapes the potential well corresponding to the native (exchange-incompetent) state and spends some time in the local minimum basin, which is comprised of various microstates representing a random coil state of the protein. If the reverse activation energy barrier (separating the local free energy minimum from the global one) is low, the protein will sample the unstructured state only transiently before returning back to the native state (Fig. 5.9*a*). As a result, only a small fraction of all labile hydrogen atoms will be exchanged upon a single unfolding event. An additional complication, which will be ignored here arises from the fact that many microstates of a random coil state can have residual protection, making certain amides unavailable to exchange. In this case, the complete exchange will require adequate sampling of all microstates within the basin.

Such a scenario, which is commonly referred to as the *EX2 exchange mechanism,* is realized if the residence time in the local minimum basin ($1/k_{cl}$) is much shorter than the characteristic time of exchange of an unprotected labile hydrogen atom ($1/k_{int}$). In this case, the probability of exchange for even a single amide during an unfolding event will be significantly $\ll 1$. The overall rate of exchange will be defined by both the frequency of unfolding events (k_{op}) and the probability of exchange during a single opening event:

$$k^{HDX} = k_{op}(k_{int}/k_{cl}) \qquad (5\text{-}3\text{-}3)$$

or, after regrouping,

$$k^{HDX} = k_{int}\,K \qquad (5\text{-}3\text{-}4)$$

where K is an effective equilibrium constant for the unfolding reaction and is determined by the free energy difference between the two states of the protein. While in NMR measurements, k^{HDX} is simply a rate of depletion of a number of protein molecules labeled with D at a specific amide, HDX MS measurements would regard this rate constant as a cumulative rate of exchange. In other words, k^{HDX} is an ensemble-averaged rate of loss of the entire D content (for all amides combined). It can be shown that under these conditions both rate constants (measured by NMR and MS) would be numerically equal, although they actually have different meanings.

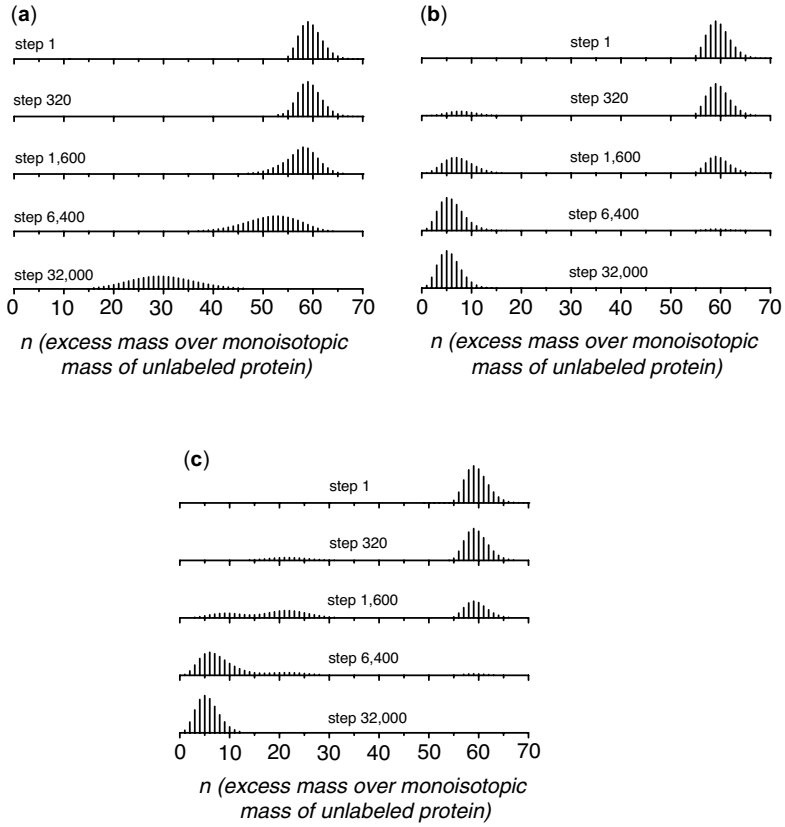


Figure 5.10. Simulated HDX MS patterns for mimimalistic models of two-state protein systems whose energy surfaces are depicted on Figure 5.9. Proteins are assumed to be fully deuterated prior to exchange (infinite dilution in a protiated buffer solution). Simulation parameters correspond to (*a*) EX2, (*b*) EX1, and (*c*) intermediate ($k_{int}/k_{cl} = 0.5$) EXX conditions.

Figure 5.10*a* shows a simulated pattern of HDX under these conditions (EX2 limit for a two-state system with 55 identical amide hydrogen atoms).

Raising the reverse-activation energy barrier separating the local minimum basin from the potential well representing the global energy minimum will decrease the refolding rate k_{cl}, leading to an increased residency time of the protein in the exchange-competent state (Fig. 5.9*b*). This, of course, will lead to an increase of the exchange probability for each labile hydrogen atom during a single unfolding event, unless the intrinsic exchange rate also decreases substantially. If the barrier is sufficiently high (so that $k_{int} >> k_{cl}$), the protein will become trapped in the unfolded state long enough to allow all labile hydrogen atoms to be exchanged during a single unfolding event. In this case (commonly referred to as the *EX1 exchange limit*) the exchange rate will be determined simply by the rate of protein unfolding:

$$k^{HDX} = k_{op} \qquad (5\text{-}3\text{-}5)$$

A pattern of HDX MS simulated under these conditions is presented in Figure 5.10*b*, which clearly shows distinct contributions from both states of the protein. In this situation the rate constant k^{HDX} defined by Eq. 5-3-5, is simply a depletion rate of the number of D-labeled protein molecules. Therefore, the physical meaning of the MS-measured exchange rate in the EX1 regime is identical to that derived from HDX NMR experiments.

Finally, in the intermediate-exchange regime (when the values of k_{int} and k_{cl} are comparable) the residency time in the exchange-competent state is long enough to have one or more protons exchanged during each unfolding event, but too short to have the entire set of all labile hydrogen atoms exchanged at once. The k^{HDX} rate constant, as measured by NMR, will reflect a gradual decrease of the protein population with a certain amide retaining its D label:

$$k^{HDX} = k_{op}(k_{int}/k_{int}+k_{cl}) \qquad (5\text{-}3\text{-}6)$$

The situation with the HDX MS measurements under the same conditions will be much more complicated, as suggested by a convoluted appearance of the exchange pattern simulated under these conditions (Fig. 5.10*c*). While the isotopic distribution is expected to exhibit a bimodal character, the distance (or mass difference) between the two clusters increases as the exchange progresses. As a result, two apparent rate constants

will be measured in a single experiment, one describing the changes in the relative abundance of the two clusters and other one related to the shift of the lower *m/z* isotopic cluster. The former rate constant is determined solely by the protein unfolding rate, as is the case under EX1 exchange conditions (Eq. 5-3-5). The second rate constant depends on both k_{int} and k_{cl}, as well as the frequency of unfolding events k_{op} in a fashion similar to (Eq. 5-3-3). In this case, the second "apparent" rate constant should be defined as a rate of shift of the monoisotopic peak in the low *m/z* cluster. Therefore, if the value of k_{int} is known, HDX MS measurements carried out under the intermediate conditions (for lack of a better term, we will use EXX throughout this chapter) may allow the values of both k_{op} and k_{cl} to be determined in a single experiment. This contrasts with measurements conducted under the EX1 and EX2 conditions, which provide information only on k_{op} (EX1 regime) or the k_{op}/k_{cl} ratio (EX2 regime).

The simplistic two-state model considered in this section is, of course, an overly idealistic representation of the protein dynamics, although it provides a good start in our discussion of how the dynamics of real protein systems are reflected in their HDX profiles. The following sections present a more detailed discussion of the HDX measurements carried out under the EX1 and EX2 exchange conditions as applied to more sophisticated systems.

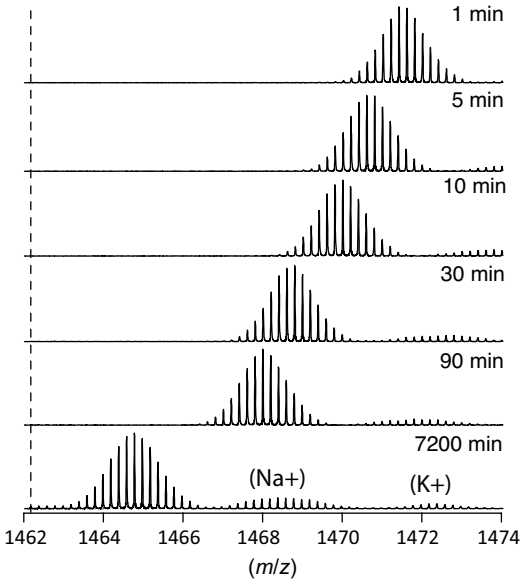


Figure 5.11. The HDX of *Chymotrypsin Inhibitor 2* under conditions favoring the EX2 exchange mechanism (the refolding rates are high, while the intrinsic exchange rates are moderate). The exchange was initiated by diluting a fully deuterated protein in a protiated solution (10 mM $CH_3CO_2NH_4$, pH adjusted to 7.0, temperature 23°C; a 1:20 dilution, *v*:*v*). The position of the exchange "end-point" is indicated with a dotted line.

5.3.3. A More Realistic Two-State Model System: Effect of Local Fluctuations on the Global Exchange Pattern Under EX2 Conditions

HDX of most proteins under near-native conditions (aqueous solutions maintained at or near neutral pH, moderate temperatures, and reasonable ionic strengths in the absence of denaturants) almost always follows EX2-type kinetics, since the refolding rates are very high. These conditions favor the natively folded protein conformation very strongly and the non-native states become populated only transiently. An example of such behavior is presented in Figure 5.11. However, EX2-type exchange can also be observed under denaturing conditions, as long as the intrinsic exchange rate is significantly lower than the refolding rate, which can be achieved, for example, under mildly acidic conditions. Analysis of the HDX MS profiles recorded under near-native conditions reveals that real proteins almost never exhibit a simple single-exponential kinetics, as could have been expected based on our consideration of a two-state model (Fig. 5.10*a*). Even in the case of two-state protein systems [e.g., chymotrypsin inhibitor 2 (CI2)], the exchange follows biphasic exponential kinetics (Fig. 5.12). The HDX MS measurements detect significantly higher initial protection compared to HDX NMR experiments carried out under similar conditions (90), as several amides (within the fast phase) exchange too fast to be measured on the time scale of a typical HDX NMR experiment. Increasing the solution temperature obviously accelerates the exchange kinetics due to increase of both k_{int} and k_{op}, while keeping it biphasic (Fig. 5.12). Further increase of the solution temperature

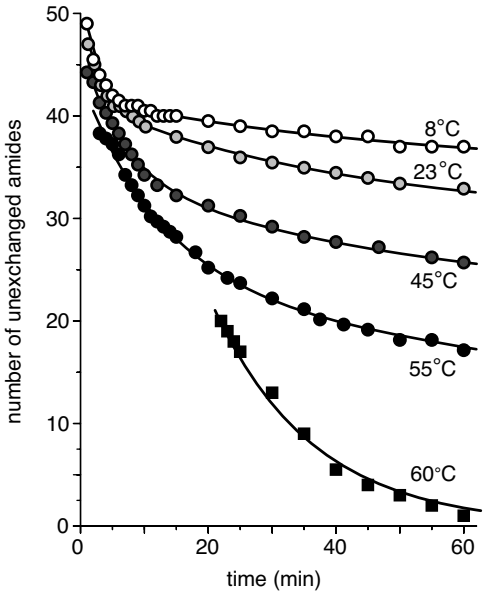


Figure 5.12. Hydrogen-exchange kinetics of CI2 measured by HDX MS at different solution temperatures under "near-native" conditions (10 mM $CH_3CO_2NH_4$, pH adjusted to 7.0). Total number of amide hydrogen atoms is 59, only 30 of which are protected on the time scale of ^{1}H NMR experiments (90).

(to 60°C) results in collapse to monoexponential HDX kinetics with a single apparent rate constant. However, in this case the fast exchange phase is simply too fast to be detected by MS without using more sophisticated time-resolved techniques. A detailed discussion of HDX MS measurements on the sub-second time scale will be presented in Chapter 6.

The most glaring deficiency of the simplistic two-state model discussed in Section 5.3.2 is that it assigns zero conformational freedom to the protein in its native state (N). As seen in Chapter 1, this assumption ignores the dynamic character of native conformations. It is now commonly accepted that native protein conformations in solution are not static structures, but rather continuously sample various microstates. Although the N $\rightleftharpoons$ U transitions discussed above are important contributors to such dynamics, they are very rare (particularly under native conditions). Much more frequent dynamic events are "local fluctuations", dynamic processes that result in transient (and localized) amide deprotection that appears to be uncooperative and denaturant-independent (91). Such amides usually exhibit fast, denaturant-independent exchange and typically reside at or near the protein surface. Therefore a solvent-penetration model can also be invoked to explain this behavior (92).

Our view of local fluctuations invokes the notion of an activated state N*, which is a collection of microstates, each having a structure (or, more precisely, amide protection pattern) identical to that of the native conformation with the exception of one (or several) amides at the protein surface. The reverse activation energy barrier separating this activated state from the ground-state $\Delta G^{\ddagger}_{N^* \rightarrow N}$ is close to zero, so that any transition from N to N* would be very short lived. The energy of this activated state, ΔG_{N^*}, is significantly lower than the activation energy for the N $\rightarrow$ U transition, so that the fluctuations occur much more frequently than the global unfolding events. In most cases, however, there would be a set of amides that would never become accessible to solvent via a local fluctuation event, hence the limited amplitude of the exchange in the fast phase. Labile hydrogen atoms that remain protected during local fluctuations could only be exchanged via a global unfolding event (in the case of a two-state protein e.g., CI2, it would be an N $\rightarrow$ U transition). It is these rare events that give rise to a slow exchange phase. Figure 5.13*a* shows a simulated HDX MS pattern of a two-state model protein under EX2 conditions that takes into account local structural fluctuations in the native state.

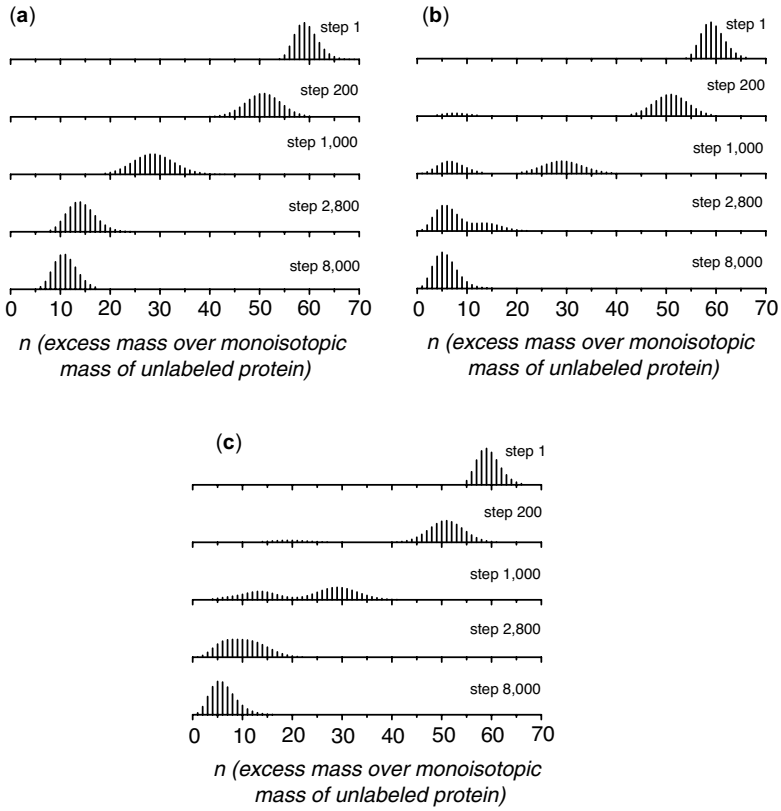


Figure 5.13. Simulated HDX MS patterns for a two-state protein model that accounts for transient local fluctuations (limited exchange competence is assigned to the native state). Proteins are assumed to be fully deuterated prior to exchange (infinite dilution in a protiated buffer solution). Simulation parameters correspond to (*a*) EX2, (*b*) EX1, and (*c*) intermediate ($k_{int}/k_{cl} = 0.5$) EXX conditions.

To allow limited conformational freedom within the native state N, we consider an activated state N^* whose structure and energetics are defined as follows. We assume N^* to be a heterogeneous state comprised of a large number of equienergetic structures, with a total number of amides exchangeable from N^* being *L*. The free energy of this state is significantly lower than that of the globally unfolded state (i.e., $\Delta G_U >> \Delta G_{N^*} >> \Delta G_N$), while the reverse activation energy barrier $\Delta G^{\ddagger}_{N^* \rightarrow N}$ is zero or very close to zero. The former feature ensures that N^* is sampled by a protein molecule much more often than the globally unfolded state. The latter feature ensures that such excursions are very short (e.g., these states are transient). Therefore, even though N^* is allowed to have certain conformational freedom, the probability of exchanging even a single amide from N^* during its lifetime would be very small. As a result, the $N \rightleftharpoons N^*$ transitions will all have characteristics of local structural fluctuations, despite the fact that N^* is formally introduced as a third state of the protein with limited exchange competence. Such presentation of local fluctuations differs from a traditional view, which assumes that every local fluctuation leads to a unique conformation (i.e., such local minima are separated from one another by significant barriers, so that these microstates cannot be grouped together). The reverse activation energy barriers separating each of these states from the exchange-incompetent state may also be significant, in which case the exchange will actually proceed through an EX1 mechanism. It is easy to show, however, that the HDX patterns produced by the two systems would be identical, and the fast phase of the exchange would still have all the characteristics of EX2 type kinetics (uncorrelated exchange). Since the exchange of all labile amide hydrogen atoms is now accomplished in two steps, broadening of the isotopic cluster, which was so evident in Figure 5.10*a*, becomes much less significant.

Local structural fluctuations in real proteins are not all identical, as the protection factors exhibit significant variation among various sites. The primary determinants are the "rigidity" of local structure (β-sheets > α-helices > Ω-loops), the degree of burial (88), as well as the presence of stabilizing tertiary contacts. In many cases, however, structural fluctuations giving rise to exchange with similar kinetic characteristics can be grouped together, giving rise to a hierarchy of local fluctuations.

5.3.4. Effects of Local Fluctuations on the Global Exchange Pattern Under EX1 and Mixed (EXX) Conditions

Increasing the lifetime of the globally unfolded state U of CI2 (e.g., by selecting mildly denaturing conditions and/or increasing the intrinsic exchange rate) leads to switching the HDX kinetics to the EX1 regime, as suggested by the *bimodal* appearance of the isotopic cluster (Fig. 5.14). Note, however, that there is an important difference between the experimentally measured HDX profile and the one simulated for a simplistic two-state system (Fig. 5.10*b*). While both isotopic clusters are static (vis-à-vis their position on the *m/z* scale) in the simulated spectra, there is clearly a noticeable shift of the higher *m/z* cluster (corresponding to the native conformation) in the experimentally obtained HDX MS profiles of CI2. It seems natural to assume that the local conformational fluctuations affect protein dynamics under these mildly denaturing conditions as well, giving rise to the partial EX2 type character of the exchange reactions. The simulated HDX MS pattern that takes into account a possibility of local structural fluctuations (Fig. 5.13*b*) is fully consistent with the experimentally measured HDX kinetics of CI2.

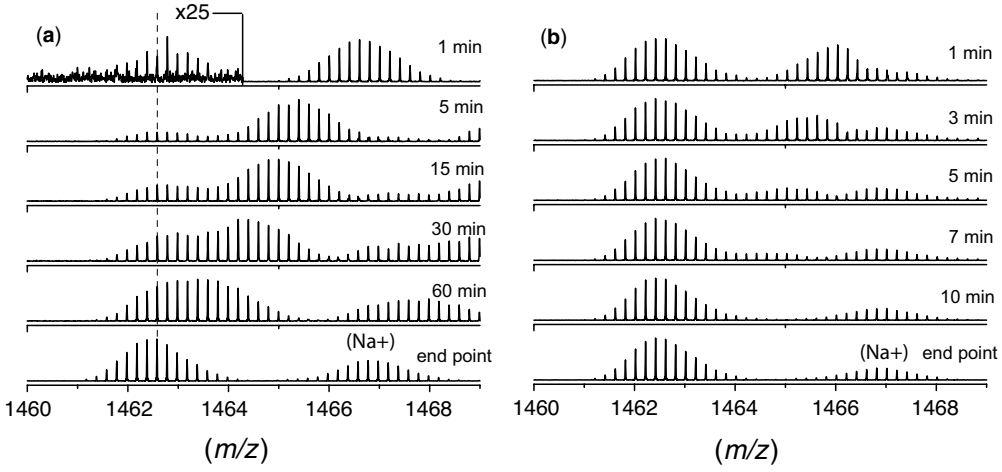


Figure 5.14. The HDX of *chymotrypsin inhibitor 2* under conditions favoring the EX1 exchange mechanism (the refolding rates are low compared to the intrinsic exchange rates). The exchange was initiated by diluting a fully deuterated protein in a protiated solution (10 mM $CH_3CO_2NH_4$, pH adjusted to 11.0 with NH_4OH, 60% CH_3OH). Solution temperature is 8°C (*a*) and 35°C (*b*).

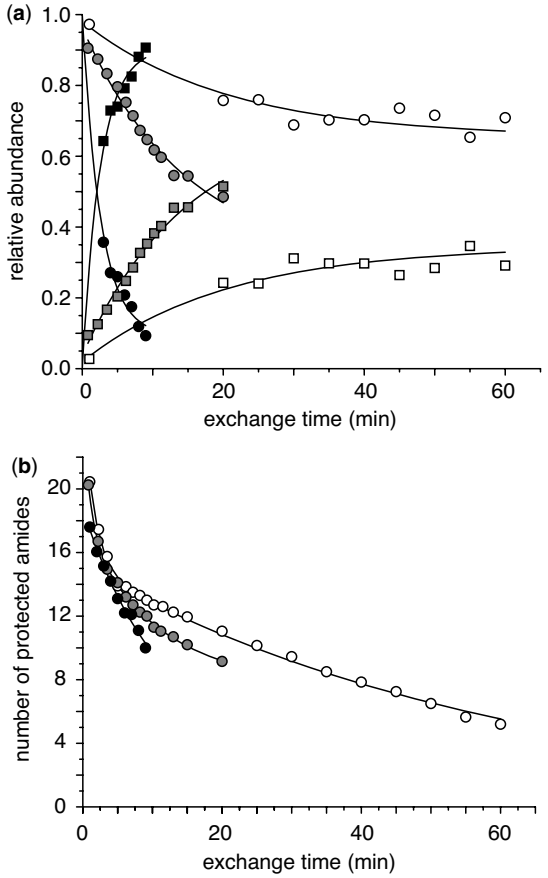


Figure 5.15. Hydrogen-exchange kinetics of CI2 measured by HDX MS at different solution temperatures (8°C, white; 23°C, gray; and 35°C, black) under conditions favoring EX1 exchange mechanism (see Fig. 5.14 for details). (*a*): Changes in relative abundance of isotopic clusters corresponding to "protected" (circles) and "fully exchanged" (squares) protein molecules. (*b*) The EX2-type kinetics reflecting local fluctuations within the "protected" conformation.

We have already seen in Eq. 5-3-5 that the rate of disappearance of the ionic signal representing the native state under EX1 conditions is in fact the rate of protein unfolding k_{op} (an N $\rightarrow$ U transition). Another quantitative characteristic of protein dynamics that can be extracted from these data relates to the local structural fluctuations within the native conformation (the rate of the gradual mass shift of the higher *m/z* cluster will be determined by Eq. 5-3-3, although k_{op} and k_{cl} in this case relate to local dynamic events, rather than to global unfolding). The kinetics of both processes is greatly influenced by solution temperature (Fig. 5.15), which may allow some useful thermodynamic data (e.g., unfolding entropy) to be extracted from such measurements.

The most complicated exchange pattern is observed in this two-state protein when the experiment is carried out under the intermediate (EXX) exchange conditions. Three distinct features are evident in both experimentally measured (Fig. 5.16*a*) and simulated (Fig. 5.13*c*) HDX MS profiles. These are the global unfolding and local fluctuations within the native state, which manifest themselves in the same ways as under the EX1 regime. Additionally, a gradual mass shift of the lower *m/z* isotopic cluster is an indicator of the EXX regime, as discussed earlier (incomplete exchange during each global unfolding event). In principle, exchange under these conditions may provide a wealth of information on protein dynamics at all levels: k_{op} (N $\rightarrow$ U) from changes in the relative abundance of the two isotopic clusters, $K^* = k_{op}{}^*/k_{cl}{}^*$ (local fluctuations within the native state) from the mass shift of the higher *m/z* isotopic cluster, and $K = k_{op}/k_{cl}$ (N $\rightarrow$ U) from the mass shift of the lower *m/z* isotopic cluster (Fig. 5.16*b–d*). A more detailed analysis of HDX MS patterns in a two-state system, as well as details of modeling can be found in (93,94).

5.3.5. Exchange in Multistate Protein Systems: Superposition of EX1 and EX2 Processes and Mixed-Exchange Kinetics

Most proteins are not simple two-state systems, but possess one or more "intermediate" states that are usually only partially structured or else their structure may exhibit a significant degree of flexibility. The presence of such intermediate states can be detected only indirectly by HDX MS under EX2 conditions (as an intermediate phase in the overall exchange kinetics, which is faster than global unfolding but slower than the local fluctuations). These intermediate states can be observed distinctly (at least, in principle) under EX1 conditions. However, this can only be achieved if the amide protection within the intermediate state is significant (to make it distinct from the fully unfolded state), but noticeably different from that of the native state. Another important requirement is that all state-to-state transitions in this system follow the EX1 mechanism. This latter requirement is not satisfied in many practically interesting cases (e.g., pH or alcohol induced). As a result, some states can escape direct detection (or else their residual protection can be misread) in a straightforward HDX MS experiment. One interesting example is presented in Figure 5.17, where the HDX of a small protein ubiquitin (Ub) is carried out under conditions known to cause a fraction of the protein molecules to assume a molten globular conformation (A-state), as well as a fully unstructured state (U) in addition to an abundant native conformation (N). However, the HDX profiles measured under such conditions are only bimodal (the fully unfolded state U escapes direct detection). The reason for such behavior is that while the N $\rightarrow$ A transition under these conditions appears to follow the EX1 exchange mechanism, sampling of the U-state occurs primarily through structural fluctuations of the A-state (direct N $\rightarrow$ U transitions are very rare), leading to an apparent EX2

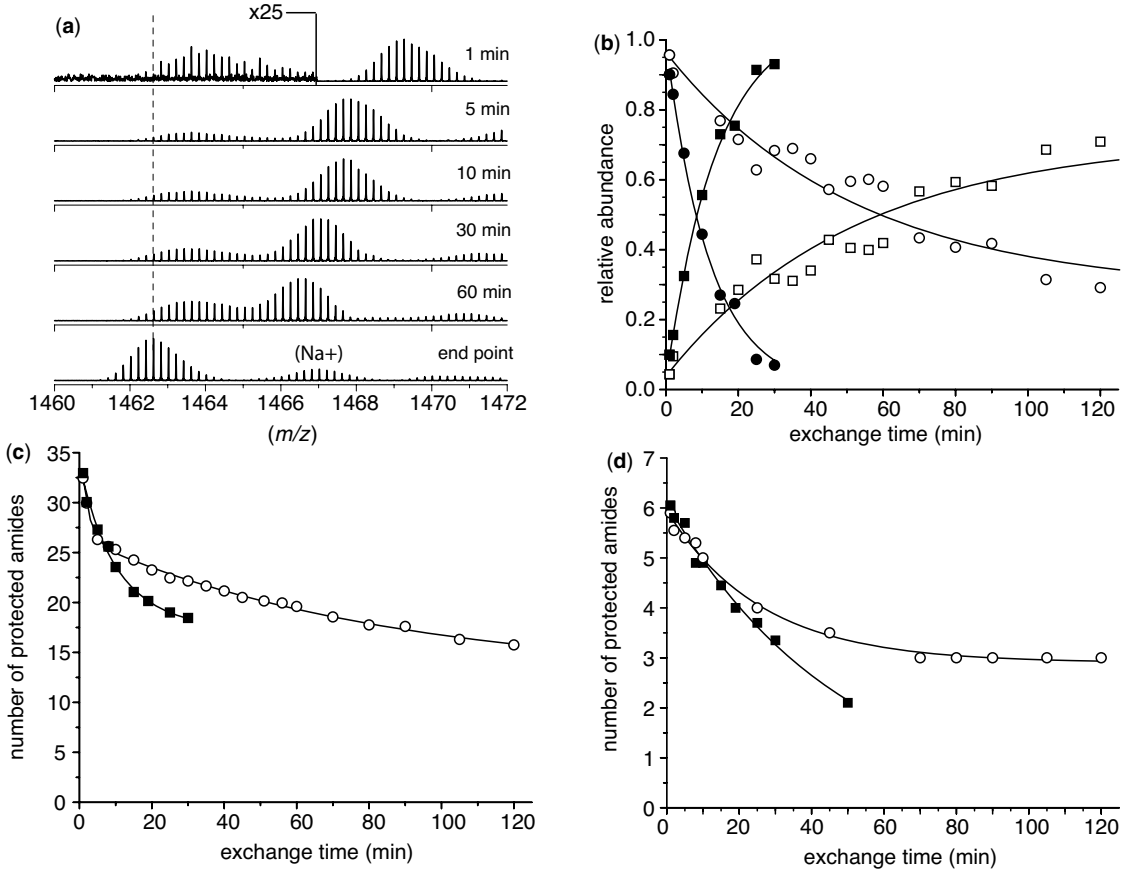


Figure 5.16. HDX of *chymotrypsin inhibitor 2* under the intermediate exchange conditions (EXX regime, the refolding rates are comparable to the intrinsic exchange rates). The exchange was initiated by diluting a fully deuterated protein in a protiated solution (10 mM $CH_3CO_2NH_4$, pH adjusted to 10.0 with NH_4OH, 70% CH_3OH). (*a*) Evolution of the isotopic distribution as a function of exchange time at solution temperature 8°C. (*b*) Changes in relative abundance of isotopic clusters corresponding to "protected" (circles) and "unprotected" (squares) protein molecules. (*c*) The EX2-type kinetics reflecting local fluctuations within the "protected" conformation. (*d*) Mass shift of the isotopic cluster corresponding to the "unprotected" protein molecules. The two data sets in panels (*b–d*) correspond to solution temperature of 8°C (white) and 23°C (black).

type exchange in the U-state. Such behavior is illustrated with a three-state energy diagram presented in Figure 5.18. The reverse activation energy barrier separating the A-state from the native conformation is sufficiently high, such that each N $\rightarrow$ A transition lasts long enough to allow full exchange of all amides unprotected in the A-state. At the same time, the barrier separating the U-state from the A-state is very low, so that the U-state becomes populated only transiently, hence the EX2 character of the exchange kinetics corresponding to the A $\rightarrow$ U transition. A more detailed analysis of HDX MS patterns in a three-state system can be found in (94,95).

Despite the limitations discussed in this section, it is possible in some cases to observe truly multimodal isotopic distributions indicative of the presence of multiple (three or more) protein conformations in solution under equilibrium (Fig. 5.19). More examples of detecting multiple (kinetic) intermediate states will be presented and discussed in Section 6.2. A final comment that should be made in connection with our discussion of global HDX patterns measured by MS relates to one particularly annoying artifact that often complicates data interpretation. As mentioned in Section 3.2.2, peptide and protein ions produced by ESI often contain one or more Na^+ and/or K^+ ions. These adduct ion peaks may overlap with and obscure the details of the convoluted bi- and multimodal isotopic distributions observed in the course of HDX MS measurements. Fortunately, high-resolution MS a offers an elegant way to solve this problem. Both Na and K have single stable isotopes with a negative mass defect. On the other hand, the mass difference between ^1H and ^2H is +1.00063 u. Therefore, Na^+ and K^+ adducts can often be resolved from their deuterium-containing isobars (see, e.g., Fig. 5.17*b*).

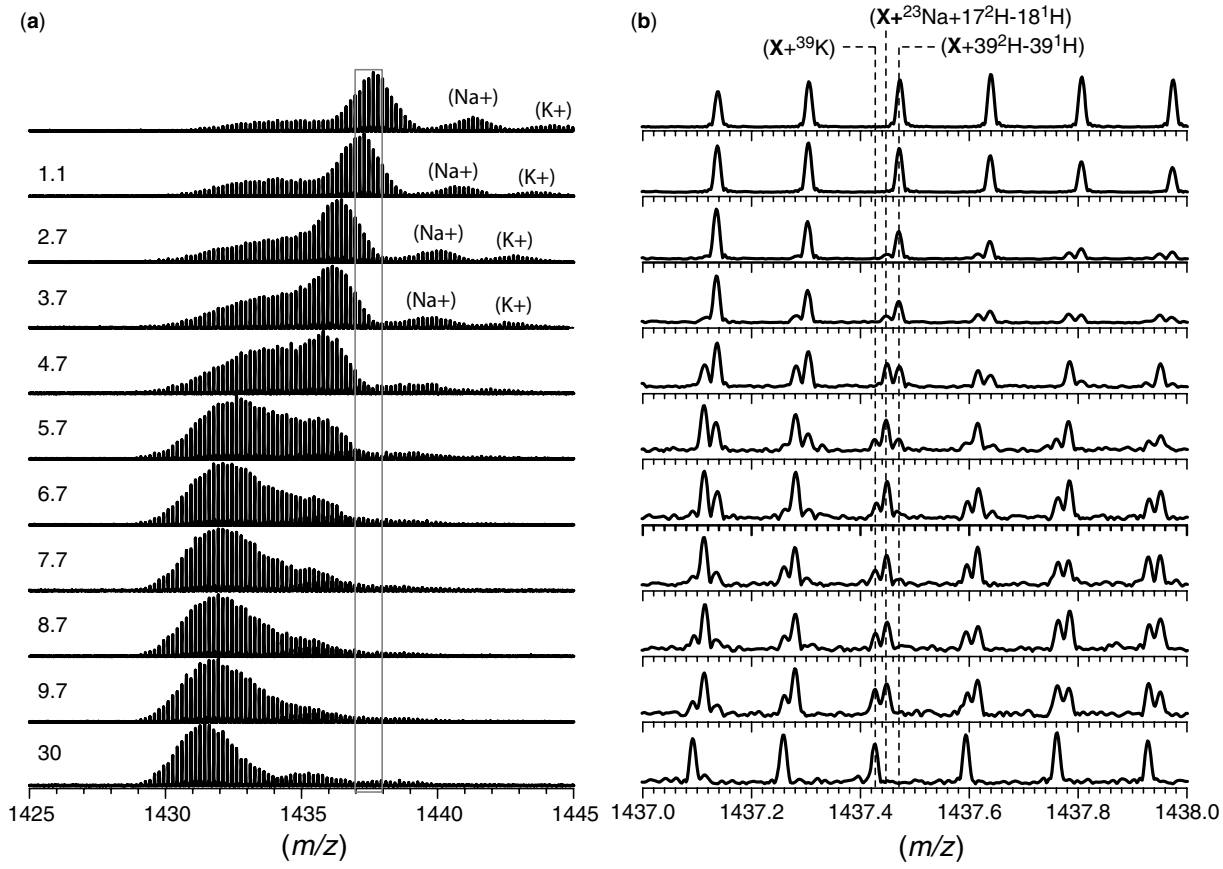


Figure 5.17. (*a*) HDX of *ubiquitin* under conditions favoring EX1 type exchange (pH 7.0, 60% CH_3OH). Only N → A transition occurs under the EX1 regime, while the complete exchange occurs via fluctuations of the A-state and has all characteristics of uncorrelated (EX2-type) exchange. (*b*) Zoomed regions of the spectra show mass-resolved isobaric peaks (protonated and alkali metal cationized ions having different ^2H content).

The ability of HDX MS to reveal the presence and characterize the behavior of distinct intermediate states is quite unique, as all other techniques used to monitor the progress of HDX reactions generate information averaged across the entire ensemble of states, thus complicating the detection of distinct conformations. Although in some instances certain information about the intermediates may be obtained by grouping amides with similar exchange kinetics into cooperative unfolding–refolding units, or *foldons* (79), this approach may result in artifacts (97,98).

5.4. MEASUREMENTS OF LOCAL PATTERNS OF HYDROGEN EXCHANGE IN THE PRESENCE OF NON-NATIVE STATES

5.4.1. Bottom-Up Approaches to Probing the Local Structure of Intermediate States

Mass spectrometry is unrivaled as far as being able to provide detailed information on the covalent structure of proteins and peptides using only minute quantities of analyte and, indeed, has become a primary experimental tool in proteomics. However, using MS to localize labile ^1H (or ^2H) atoms within a protein or a polypeptide is not a trivial task. Proteolytic fragmentation and separation of the fragments prior to MS analysis will inevitably alter both the overall isotopic content and its distribution across the polypeptide chain, even if carried out for short periods of time. This problem can be at least partially remedied for one particularly important class of labile hydrogen atoms, those belonging to the backbone amide groups. It was realized a quarter of a century ago that one ubiquitous proteolytic enzyme, pepsin, is active within the pH range 2.5–3, under which condition intrinsic amide exchange rates are typically minimal, which makes it suitable for probing amide protection (99,100). All side-chain labile hydrogen atoms exchange fast in this pH range (Fig. 5.8), and the commonly used terms *quenched exchange* or *slow exchange* refer only to backbone amides. Even the amide hydrogen atoms continue to exchange under such slow exchange conditions,

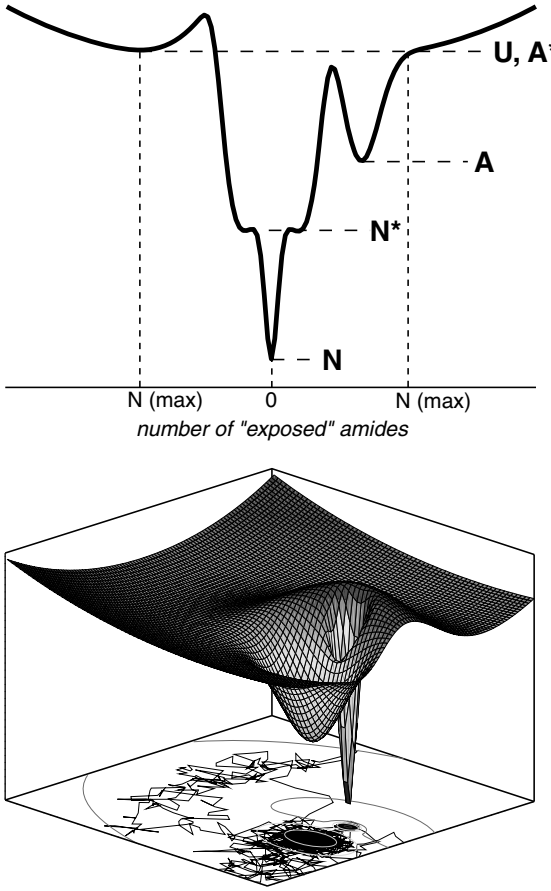


Figure 5.18. A minimalistic representation of a three-state protein system. Transition from the N-state (global energy minimum) to a partially structured molten globule-like A-state occurs under EX1 conditions. Transition from the A-state to a fully unstructured U-state occurs under EX2 conditions.

and the local exchange details can be maintained only if the sample handling is relatively fast and is performed at low solution temperature (typically 0–4°C).

Nearly two decades ago, Zhang and Smith (101) complemented HDX with peptic digest of the protein carried out under slow exchange conditions, followed by mass analysis. The proteolytic fragmentation of the protein was carried out by incubating it in pepsin solution at 0°C for 10 min, followed by mass analysis of peptic fragments using MS. This procedure is still frequently used today [with the exception of fast-atom bombardment (FAB) MS initially employed by Smith being now almost universally substituted with LC–ESI MS], although the large quantity of pepsin required for rapid digestion and its autolysis products can interfere with the chromatographic step, adversely affect the ESI process and complicate data analysis (102). These problems can be remedied by utilizing an online approach (102,103), which makes use of an immobilized pepsin column in tandem with an HPLC column to carry out fast and efficient digestion and desalting–preconcentration– separation of peptic fragments prior to their mass analysis (Fig. 5.20). Mammalian pepsin remains the most popular proteolytic enzyme suitable for HDX MS work, although the search for a suitable alternatives continues (104). Perhaps the most successful complement to pepsin in HDX MS work reported to date are proteases from *Aspergillus satoi* (Type XIII) and *Rhizopus* sp. (Type XVIII) (105). Like pepsin, neither of these proteases is specific, but there is good reproducibility in the protein digestion patterns. The search for other acidic proteases that could be used in HDX MS work continues (106,107).

In some cases, the separation step can be bypassed, as utilization of high-resolution mass analyzers allows different peptic fragments to be resolved even if their peaks in the mass spectra partially overlap. This approach, however, is not frequently used due to the concerns that elimination of peptic fragment separation prior to MS measurements will result in significant signal suppression and reduce the number of peptides that can be identified and subsequently used in the analysis of HDX MS data.

An example of using local HDX MS measurements to detect and characterize non-native protein states is shown in Figure 5.21, where it is used to investigate partial unfolding of interferon-β1a triggered by alkylation of its single free cysteine residue (30). Earlier in this chapter, we saw that the partial loss of structure is readily revealed by the analysis of ionic charge-state distributions of intact and alkylated forms of this protein under native conditions (see Fig. 5.2), although it does not provide any information that would allow the specific unfolding events to be localized within the protein structure. This gap is easily filled by local HDX MS, which provides information on backbone protection patterns for both forms of this protein (30). The HDX MS identifies several segments in the alkylated form whose stability is greatly compromised by alkylation. One example is shown in Figure 5.21, where the evolution of the isotopic distribution of a peptic fragment [L88 – L102] is traced over the first 120 s. of exchange in solution. Very slow uptake of deuterium is exhibited by the fragment derived from intact IFN, while protein alkylation results in a dramatic acceleration of the exchange kinetics of this segment. The presence of several such segments whose exchange kinetics is significantly altered by the alkylation event, clearly indicates a change in conformation and/or dynamics induced by the alkylation. The detailed information on the backbone protection of intact and modified IFN provided by HDX MS measurements allows the conformational properties of the two forms of the protein to be compared directly (30,32). Perhaps the most important conclusion derived from the HDX MS work is the dramatic destabilization of one of the helices in the partially unstructured IFN (see the structure in Fig. 5.2*b*). As we will see in Chapter 7, such observations allow predictions to be made regarding the influence exerted by the structural changes on the protein function.

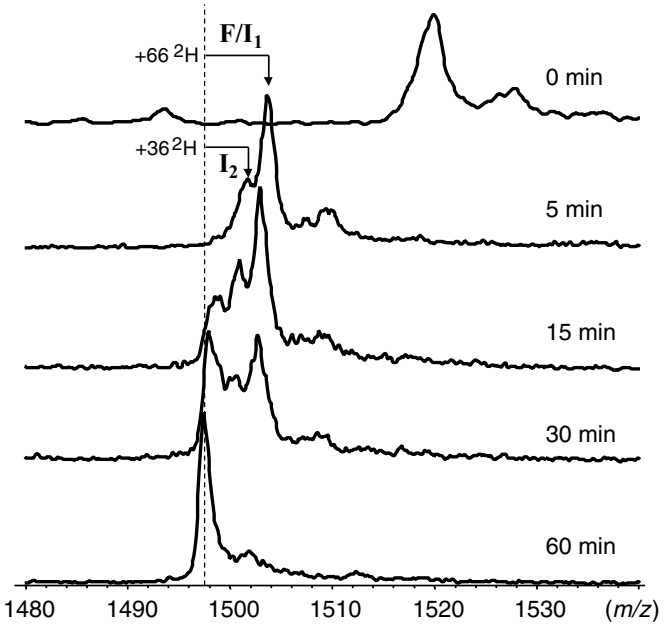


Figure 5.19. The HDX of *pseudo-wild type* cellular retinoic acid binding protein I under conditions favoring EX1 exchange regime (acid-catalyzed exchange). Time evolution of a *12+* ion peak profile throughout the course of the exchange reaction clearly shows presence of at least three states under equilibrium differing by the degree of backbone protection. The dashed line on the left indicates the position of the fully exchanged protein ion peak. Reproduced with permission from (96).

The data presented in Figure 5.21 are noteworthy for another reason. The light-gray trace shown on the background represents the experimental end-point of the exchange reaction for peptic fragment [L88–L102], that is, isotopic distribution of this peptide derived from the fully denatured protein that was exposed to the exchange buffer for several hours to allow for complete exchange of all labile hydrogen atoms, and then subjected to the same experimental procedure as the proteins undergoing HDX for fixed periods of time (represented by other traces in that same figure). Under these conditions, all of the backbone amides should have been exchanged. To be, more precise, the fraction of retained ^{1}H atoms should match the proportion of these isotopes in the exchange buffer, which ranges

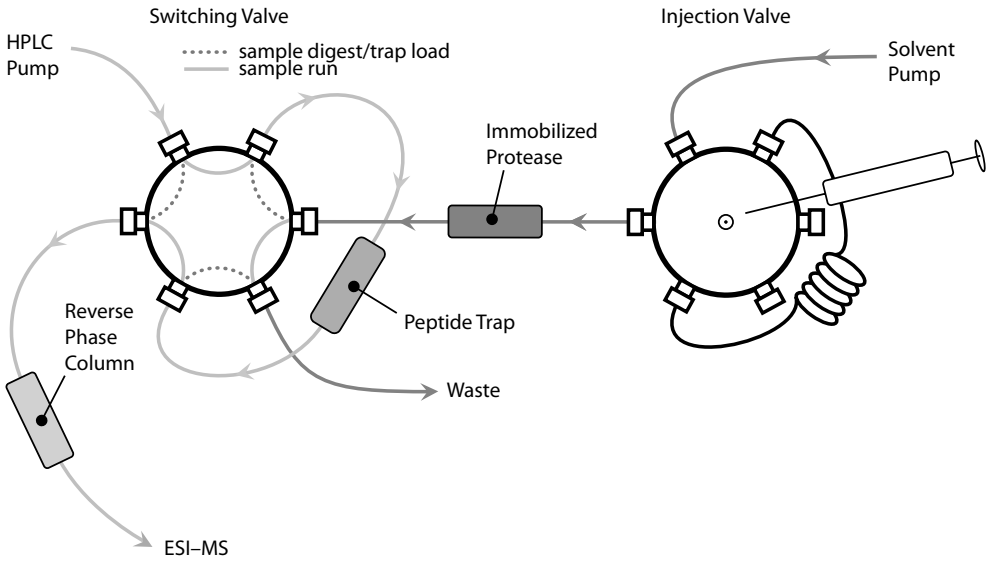


Figure 5.20. Schematic representation of HDX MS set-up utilizing online proteolysis and sample clean-up prior to MS analysis.

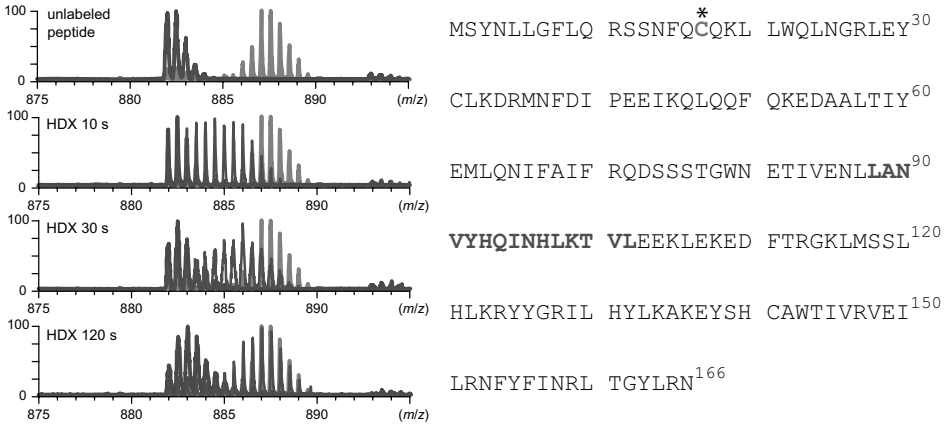


Figure 5.21. Evolution of isotopic distributions of peptic fragments (88–102) derived from intact (dark gray) and NEM alkylated (black) interferon-β1a throughout the course of HDX. The endpoint of the exchange reaction is indicated with a light gray trace (isotopic distribution of a fully exchanged peptide). Location of this peptide within the amino acid sequence of IFN is shown on the right. [Adapted with permission from (32). Copyright © 2010 American Society for Mass Spectrometry.]

between 1 and 10% depending on the specific D_2O/H_2O dilution ratio used in the experiment. However, an attentive reader will notice that the actual level of deuteration of the backbone amides in this peptide is noticeably $< 90\%$; in fact, the position of the light-gray isotopic distribution on the *m/z* scale indicates inclusion of only 11 ^2H atoms in the peptide having 14 backbone amides. The lower than expected apparent level of deuteration of peptic fragments at the endpoint of the exchange reaction is caused by the relatively inefficient exchange of the completely labeled peptide with the solvent during sample processing under slow exchange conditions. This phenomenon, known as *back-exchange*, may be accelerated during various stages of protein processing, [e.g., during the chromatographic step (108)].

Although back-exchange was frequently evaluated in early HDX MS studies using unstructured model peptides, the utility of this procedure is questionable, since the intrinsic exchange rates are highly sequence-dependent. In many instances, back-exchange may be estimated using algorithms based on context-specific kinetics data (109).* In many cases, it can be determined experimentally for each proteolytic fragment by processing a fully labeled protein using a series of steps that precisely reproduce those used in HDX MS measurements, as shown in the example considered in the previous paragraph. Typical back-exchange levels reported in the recent literature range from 10 to 50%, although significantly higher numbers have also been reported. Even if back-exchange can be accounted for, it nonetheless has a detrimental influence on the quality of HDX MS measurements by reducing the available dynamic range.

* Available at http://hx2.med.upenn.edu/download.html.

Two other factors that must be considered in evaluating the quality of bottom-up HDX MS measurements are the extent of sequence coverage and the spatial resolution. The first relates to the fact that in many cases it is impossible to have a complete coverage of the protein sequence with peptic fragments. This could be caused by a variety of factors, most notable of which is probably the suboptimal efficiency of proteolysis under the conditions that minimize back-exchange. Ironically, very high efficiency may also be detrimental, as it would generate very small peptide fragments that would be eluted with the solvent front in LC and evade detection by MS, or that may lack unique sequence information. A structural feature whose presence in proteins has a negative impact on the spatial resolution achievable in HDX MS measurements is disulfide bonding. Apart from limiting the efficiency of proteolysis, it also prevents physical separation of peptic fragments connected by the cystine–cystine bridges even if the enzymatic reaction is successful. Common disulfide-reducing agents, (e.g., dithiothreitol) are inactivated at acidic pH and, therefore, cannot be used under slow exchange conditions. The task of reducing disulfides under such conditions can be successfully carried out in many cases by an alternative reagent, tris (2-carboxyethyl)phosphine (TCEP), which retains its disulfide-reducing capacity under slow exchange conditions (110). Complete reduction of disulfide bonds may be difficult to achieve even with TCEP, particularly in disulfide-rich proteins, although certain changes in experimental protocol can alleviate this problem, at least for smaller proteins (111). Another structural feature that dramatically reduces protein susceptibility to proteolysis in HDX MS experiments is high aggregation propensity. These types of proteins sometimes can be dealt with using high-pressure digestion systems (112).

Less than desirable sequence coverage in HDX MS is related to the nonspecificity of pepsin under slow-exchange conditions. This prevents prediction a priori of the expected cleavage sites, making the task of identifying all peptic fragments more difficult. Typically, this is not a big problem for smaller proteins, where most peptide fragments can be confidently identified by mass using various search engines. However, certain structural features, such as extensive glycosylation and/or the presence of unreduced disulfide bonds, as well as large protein size, may complicate the task of identifying proteolytic fragments by dramatically expanding the sheer number of potential isobaric fragments that can be derived from the protein in question. Tandem mass spectrometry and high-resolution mass measurements become invaluable in many such cases.

The second factor that often limits the quality and usefulness of HDX MS data is inadequate spatial resolution. Longer peptic fragments may span several distinct structural regions, and the protection level averaged across the entire peptide length (as reported by HDX MS) may be meaningless. In general, a large number of fragments, particularly overlapping ones, would lead to greater spatial resolution, and hence more precise localization of the structural regions that have undergone exchange. In some favorable cases, spatial resolution in HDX MS of small proteins ($<$15 kDa) may be enhanced up to the single residue level by analyzing deuterium content of a set of overlapping proteolytic fragments (113). Single-residue resolution has yet to be demonstrated in HDX MS studies of larger proteins, although overlapping peptic fragments frequently provide at least moderate improvement of spatial resolution.

Potentially, spatial resolution can be further improved by introducing a gas phase fragmentation step during mass analysis (e.g., MS/MS). This approach was initially tested by Smith and co-workers (114) and Deinzer and co-workers (115), who implemented HDX MS/MS schemes using collisional activation (CID) of peptic fragments in the gas phase. These studies, however, yielded mixed results due to apparent scrambling in some (but not all) fragment ions (see Section 5.4.2 for a more detailed discussion of hydrogen scrambling). Later reports showed even more extensive scrambling in small peptide ions subjected to collisional activation (116), an obvious anathema to the proposed marriage of CID and bottom-up HDX MS. Nonetheless, the continuing search for a scrambling-free solution to this problem has yielded very encouraging results, with electron-based ion dissociation methods showing minimal scrambling when applied to short peptides under carefully controlled conditions (117,118). The feasibility of supplementing proteolytic fragmentation in solution with electron transfer dissociation (ETD) in the gas phase was recently demonstrated using a small model polypeptide (119). At the time of writing, several groups are actively involved in developing various HDX MS/MS schemes using ETD or electron capture dissociation (ECD) on a variety of commercially available MS platforms. It is not inconceivable that by the time this book goes to press, we will have a tool to produce truly high-resolution HDX MS data for proteins that are beyond the reach of high-resolution NMR.

In conclusion, note that although ESI MS is the most popular detection technique used in HDX MS measurements, MALDI MS can also be employed. Komives and coworkers (120) first demonstrated that the standard HDX MS protocol (exchange-quench-proteolysis) can be implemented using MALDI MS to monitor deuterium incorporation in a site-specific fashion (120). Only minor modifications of established MALDI sample preparation protocols are required to carry out such experiments. Quenching the amide HDX and peptic digest are followed by chilled, rapid drying of the samples mixed with a matrix solution whose pH is adjusted to 2.5 on the MALDI target to ensure minimal loss of deuterons from peptide amide groups. Use of MALDI offers superior sensitivity (at subpicomolar level) and eliminates the need for preconcentration and separation steps prior to MS analysis of the deuterium content of peptic fragments. This, however, come at a price, since the presence of a large number of peptic fragments in a single spot is likely to lead to signal suppression and, subsequently, loss of sequence coverage.

5.4.2. Top-Down Approaches to Probing the Local Structure of Intermediate States

A method to probe HDX kinetics locally with ESI MS without the need for proteolytic fragmentation was introduced by Anderegg et al. (121) nearly two decades ago. The method relied on the ability of mass spectrometers to produce a wealth of structural information in MS/MS experiments. In this scheme, CID was employed to fragment relatively short polypeptides ($<$ 4 kDa) in the gas phase in order to obtain residue-specific information on the location of labile hydrogen atoms. This initial report was followed by several others (122–125), where a similar approach was used to obtain protection patterns for short polypeptides. Adaptation of this approach for protein studies only became possible (96) following improvements in Fourier transform (FT) MS and hybrid time-of-flight (TOF) analyzers in the late 1990s, which resulted in a dramatic expansion of the range of proteins for which rich gas-phase fragmentation patterns could be obtained.

A schematic diagram of a typical top-down HDX MS experimental workflow is shown in Figure 5.22. The protein solution is infused directly into the ESI source, where intact protein ions are produced. The infusion may or may not be preceded by quenching the exchange reaction, depending on the specific experimental setup (e.g., if the exchange is fast, quenching may be needed in order to avoid changing the protein deuteration level during the time course of the measurement). Dissociation of the protein ions in the gas

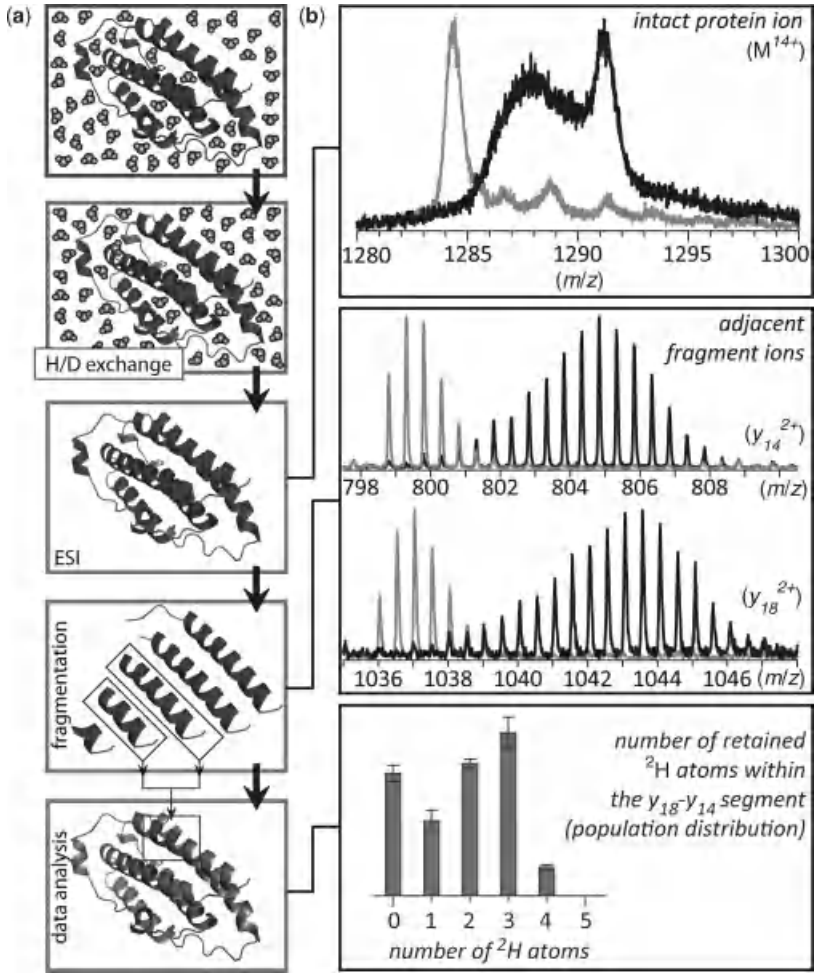


Figure 5.22. A flow-chart diagram of the top-down HDX MS (HDX MS/MS) measurements (*a*) and the steps involved in data processing (*b*). The data represent HDX MS/MS measurements of *pseudo-wild-type(wt*)* cellular retinoic acid binding protein I under mildly denaturing conditions. The existence of multiple protein conformers is evident from the convoluted appearance of the isotopic distribution of the intact protein ion). Protection in the C-terminal protein segments can be deduced from the isotopic distributions of corresponding *y*-fragments (gray traces show the endpoints of the exchange reaction). Adjacent fragment ions are used to determine deuterium content within shorter protein segments. Adapted with permission from (127) Copyright © 2009 American Chemical Society and (128) Copyright © 2009 American Society for Mass Spectrometry.]

phase will generate a series of fragment ions whose mass would reflect the number of D atoms contained within the corresponding protein segments. This is best done by comparing the fragment ion mass at a certain point of HDX reaction to that at the exchange reaction endpoint. Average masses of adjacent fragment ions (e.g., y_{14}^{2+} and y_{18}^{2+} in Fig. 5.22) would then provide an average deuteration level for a short protein segment by a simple subtraction (y_{18} - y_{14}). In a situation where the isotopic distributions of fragment ions are not symmetrical, a more sophisticated deconvolution procedure should be used, which provides an ensemble distribution of the number of D atoms within the segment in question across the entire population of protein molecules (128).

One unique advantage of top-down HDX MS measurements that cannot be matched by the classic bottom-up type experiments is the ability to obtain protection patterns in a conformer-specific fashion, which is illustrated in Figure 5.23. The isotopic profile of a fully deuterated 18 kDa protein *wt**-CRABP I is recorded following its brief exposure to the ^{1}H-based exchange buffer, and the bimodal appearance of the isotopic distribution of the molecular ion (top trace in Fig. 5.23*a*) clearly indicates the presence of at least two conformers with different levels of backbone protection. Collisional activation of the entire protein ion population generates a set of fragment ions with convoluted isotopic distributions (top trace in Fig. 5.23*b*). However, mass selection of precursor ions with a specific level of

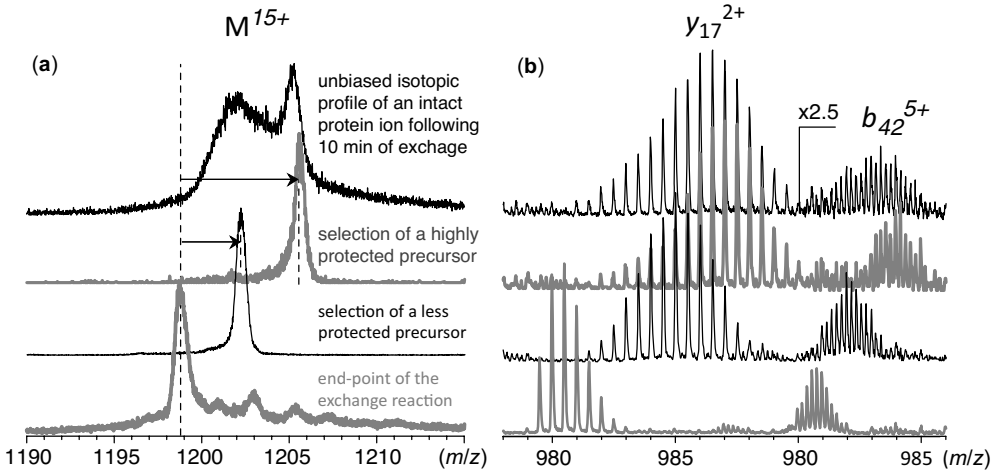


Figure 5.23. Characterization of local dynamics in *wt**-CRABP I in a conformer-specific fashion using top-down HDX MS (fully deuterated protein was exposed to $^1H_2O/CH_3CO_2N^1H_4$ at pH 3.1 for 10 min; the gray trace at the bottom corresponds to HDX end-point). (*a*) Mass selection of precursor ions for subsequent CAD (from top to bottom): broad-band selection of the entire ionic population (not conformer-specific); highly protected conformers; narrow population of less protected conformers; HDX endpoint. (*b*) Isotopic distributions of two representative fragment ions generated by CID of precursor ions shown in *a*. Selection of different ion populations as precursor ions for subsequent fragmentation was achieved by varying the width of a mass selection window of a quadrupole filter (*Q*) in a hybrid quadrupole/TOF mass spectrometer (Qq TOF MS). [Adapted with permission from (127). Copyright © 2009 American Chemical Society.]

deuterium content allows the top-down HDX MS measurements to be carried out in a conformation-specific fashion, taking full advantage of the ability of HDX MS to detect distinct conformers. For example, selective fragmentation of protein ions representing a highly protected conformation is achieved by mass-selecting a narrow population of intact protein ions with a high level of retained deuterium (the second trace in Fig. 5.23*a*). Mass-selection and subsequent fragmentation of a narrow population of protein ions with significantly lower deuterium content (the third trace in Fig. 23*a*) generates a set of fragment ions whose isotopic distributions provide information on backbone protection within non-native protein states. For example, the data presented in Figure 5.23*b* clearly indicate that the C-terminal segment of the protein represented by the y_{17}^{2+} ions retains significant structure even within the partially unfolded conformers: The amount of retained deuterium atoms reduces by only 30% as a result of switching from the precursor ion of highly protected (second trace) to less protected (third trace) populations. At the same time, selection of the precursor ion has a much more dramatic effect on the protection levels exhibited by the N-terminal segment (represented by the b_{42}^{5+} ion), where more than a twofold decrease in the amount of retained deuterium atoms is observed. Extending this analysis to other protein fragments may allow detailed backbone protection maps to be created for each protein conformer. A similar approach recently has been used by Konermann and co-workers to study the solution-phase structure of Aβ(1–42) amyloid peptide in a conformer-specific fashion (129). In addition to equilibrium intermediates, top-down HDX MS can also be used to study transiently populated intermediate states (130).

Despite the great promise of top-down HDX MS demonstrated by the studies discussed above, applications of this technique have been limited so far due to concerns over the possibility of hydrogen scrambling accompanying dissociation of protein ions in the gas phase. Indeed, several reports have pointed out that proton mobility in the gas phase may, under certain conditions, influence the outcome of top-down HDX MS measurements when collisional activation is employed to fragment protein ions (131,132). The occurrence (or absence) of hydrogen scrambling in the gas phase can be reliably detected by using built-in scrambling indicators. One particularly convenient indicator is a His-tag, a histidine-rich segment appended to wild-type sequences to facilitate protein purification on metal affinity columns. Such segments are fully unstructured in solution and, therefore, should lack any backbone protection (133). Alternatively, intrinsic scrambling indicators (e.g., internal flexible loops (132)), as well as other approaches (131) can be used to detect occurrence of scrambling. The available experimental evidence suggests that slow protein ion activation [e.g., sustained off-resonance radiation collisionally activated dissociation (SORI CAD)] always leads to hydrogen scrambling, while fast activation allows it to be minimized

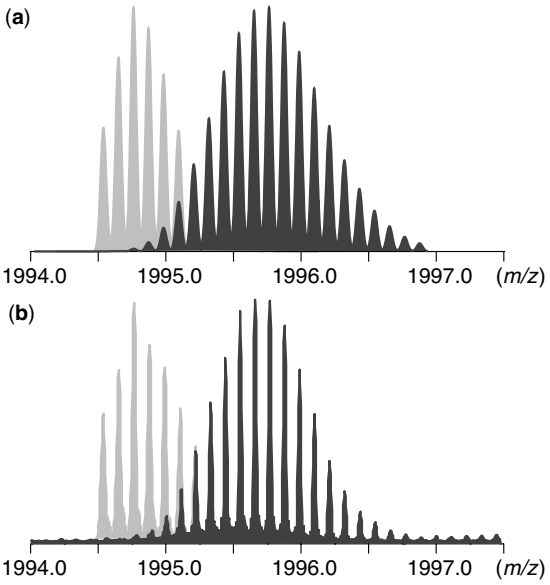


Figure 5.24. The effect of ^{13}C depletion on the isotopic distribution of *pseudo-wild-type* cellular retinoic acid binding protein I ions. Calculated (*a*) and experimentally measured (*b*) isotopic distributions corresponding to intact ions (charge state 9+) of proteins expressed in *Escherichia. coli* in growth media containing regular (dark) and ^{12}C-enriched (99.9%) glucose (light gray). [Adapted with permission from (140). Copyright © 2004 Nova Science Publishers, Inc.]

or eliminated in top-down HDX MS experiments (132). Selection of highly charged precursor ions for CID has also been shown to either eliminate or minimize hydrogen scrambling (134).

Some time ago we suggested that the specter of hydrogen scrambling in top-down HDX MS measurements may be alleviated by using electron-based fragmentation processes instead of collisional activation (135). Several recent studies demonstrated that the extent of scrambling is indeed negligible when ECD (136) or ETD (137) is used as a means to obtain fragment ions in top-down HDX MS experiments. In addition to allowing scrambling to be easily eliminated in top-down HDX MS experiments, both ECD and ETD appear to be superior to CID in terms of sequence coverage, at least for proteins in the 20 kDa range. Unlike CID, protein backbone cleavage in ECD and ETD is less specific (138) leading to a higher number of fragment ions. This translates not only to improved sequence coverage, but also to enhanced spatial resolution. In fact, in some cases it becomes possible to generate patterns of deuterium distribution across the protein backbone down to the single-residue level (136,139), whereas our earlier attempts to improve sequence coverage in top-down HDX MS experiments using CID by employing multistage fragmentation (MS^n) were unsuccessful due to massive hydrogen scrambling exhibited by the second generation of fragment ions (134).

5.4.3. Further Modifications and Improvements of HDX MS in Conformationally Heterogeneous Systems

One other difficulty facing HDX MS measurements in general (both "global" and "local" measurements) is that even under EX1 exchange conditions, the isotopic clusters corresponding to different intermediate states can overlap. This overlap would be particularly significant if the extent of protection does not differ greatly among the intermediate states involved. Structural fluctuations within these intermediate states would result in a partial EX2-like exchange component of the observed HDX pattern (see above), thus broadening the isotopic clusters and further exacerbating the problem. Even without any exchange the isotopic clusters of protein ions have substantial width due to the contributions of naturally occurring "minor" isotopes of carbon, nitrogen, oxygen, and sulfur (i.e., ^{13}C, ^{15}N, ^{18}O, ^{33}S, and ^{34}S), while the contribution from naturally occurring ^{2}H is negligible and can be safely ignored despite the large number of hydrogen atoms in a typical protein. The problem of isotopic broadening can be addressed by removing the background isotopic distributions. Depletion of ^{13}C alone is expected to give a narrower isotopic cluster for protein ions even of modest size (Fig. 5.24*a*). This can be achieved by expressing the proteins in cells grown on isotopically depleted media. An example of the isotopic distribution for an 18 kDa protein grown on 99.9% ^{12}C glucose is shown in Figure 5.24*b*. In addition to collapsing the broad natural distribution to a few isotopic peaks, one can also clearly see the monoisotopic peak of the intact protein ion. This allows the interpretation of the mass spectrum to be dramatically simplified (141), and should enable more precise local measurements of deuterium occupancies in the HDX MS experiments that utilize either proteolytic or gas-phase protein fragmentation prior to mass analysis.

Alternatively, the convoluted isotopic distributions in some cases can be decomposed to yield contributions from each protein conformation (142). Perhaps it will be possible in the future to streamline such deconvolution by using various chemometric techniques. Another important modification of HDX MS measurements that greatly improves the temporal resolution afforded by the technique utilizes a rapid mixing apparatus (143) interfaced to the ESI source and will be discussed in Chapter 6.

REFERENCES

1. Chowdhury, S. K., Katta, V., and Chait, B. T. (1990) Probing conformational changes in proteins by mass spectrometry, *J. Am. Chem. Soc.* 112, 9012–9013.
2. Loo, J. A., Loo, R. R., Udseth, H. R., Edmonds, C. G., and Smith, R. D. (1991) Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry, *Rapid Commun Mass Spectrom* 5, 101–105.

3. Konermann, L. and Douglas, D. J. (1997) Acid-induced unfolding of cytochrome *c* at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure, *Biochemistry* 36, 12296–12302.
4. Yu, X., Wojciechowski, M., and Fenselau, C. (1993) Assessment of metals in reconstituted metallothioneins by electrospray mass spectrometry, *Anal. Chem.* 65, 1355–1359.
5. Kashiwagi, T., Yamada, N., Hirayama, K., Suzuki, C., Kashiwagi, Y., Tsuchiya, F., Arata, Y., Kunishima, N., and Morikawa, K. (2000) An electrospray-ionization mass spectrometry analysis of the pH-dependent dissociation and denaturation processes of a heterodimeric protein, *J. Am. Soc. Mass Spectrom.* 11, 54–61.
6. Hanefeld, U., Stranzl, G., Straathof, A. J., Heijnen, J. J., Bergmann, A., Mittelbach, R., Glatter, O., and Kratky, C. (2001) Electrospray ionization mass spectrometry, circular dichroism and SAXS studies of the (*S*)-hydroxynitrile lyase from Hevea brasiliensis, *Biochim. Biophys. Acta* 1544, 133–142.
7. Nabuchi, Y., Murao, N., Asoh, Y., and Takayama, M. (2007) Probing the unfolding and refolding processes of carbonic anhydrase 2 using electrospray ionization mass spectrometry combined with pH jump, *Anal. Chem.* 79, 8342–8349.
8. Borysik, A. J. H., Radford, S. E., and Ashcroft, A. E. (2004) Co-populated conformational ensembles of β_2-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry, *J. Biol. Chem.* 279, 27069–27077.
9. Invernizzi, G. and Grandori, R. (2007) Detection of the equilibrium folding intermediate of β-lactoglobulin in the presence of trifluoroethanol by mass spectrometry, *Rapid Commun. Mass Spectrom.* 21, 1049–1052.
10. Kintzer, A. F., Sterling, H. J., Tang, I. I., Abdul-Gader, A., Miles, A. J., Wallace, B. A., Williams, E. R., and Krantz, B. A. (2010) Role of the protective antigen octamer in the molecular mechanism of anthrax lethal toxin stabilization in plasma, *J. Mol. Biol.* 399, 741–758.
11. Wang, G., Abzalimov, R. R., and Kaltashov, I. A. (2011) Direct monitoring of heat-stressed biopolymers with temperature-controlled electrospray ionization mass spectrometry, *Anal. Chem.* 83, 2870–2876.
12. Low, L. Y., Hernandez, H., Robinson, C. V., O'Brien, R., Grossmann, J. G., Ladbury, J. E., and Luisi, B. (2002) Metal-dependent folding and stability of nuclear hormone receptor DNA- binding domains, *J. Mol. Biol.* 319, 87–106.
13. Gumerov, D. R. and Kaltashov, I. A. (2001) Dynamics of iron release from transferrin N-lobe studied by electrospray ionization mass spectrometry, *Anal. Chem.* 73, 2565–2570.
14. van den Bremer, E. T., Jiskoot, W., James, R., Moore, G. R., Kleanthous, C., Heck, A. J., and Maier, C. S. (2002) Probing metal ion binding and conformational properties of the colicin E9 endonuclease by electrospray ionization time-of-flight mass spectrometry, *Protein Sci.* 11, 1738–1752.
15. Merrifield, M. E., Huang, Z., Kille, P., and Stillman, M. J. (2002) Copper speciation in the α and β domains of recombinant human metallothionein by electrospray ionization mass spectrometry, *J. Inorg. Biochem.* 88, 153–172.
16. Gumerov, D. R., Mason, A. B., and Kaltashov, I. A. (2003) Interlobe communication in human serum transferrin: metal binding and conformational dynamics investigated by electrospray ionization mass spectrometry, *Biochemistry* 42, 5421–5428.
17. Natalello, A., Benetti, F., Doglia, S. M., Legname, G., and Grandori, R. (2011) Compact conformations of α-synuclein induced by alcohols and copper, *Proteins* 79, 611–621.
18. Ngu, T. T., Krecisz, S., and Stillman, M. J. (2010) Bismuth binding studies to the human metallothionein using electrospray mass spectrometry, *Biochem. Biophys. Res. Commun.* 396, 206–212.
19. Kaluarachchi, H., Sutherland, D. E. K., Young, A., Pickering, I. J., Stillman, M. J., and Zamble, D. B. (2009) The Ni(II)-binding properties of the metallochaperone SlyD, *J. Am. Chem. Soc.* 131, 18489–18500.
20. Xiao, H., Kaltashov, I. A., and Eyles, S. J. (2003) Indirect assessment of small hydrophobic ligand binding to a model protein using a combination of ESI MS and HDX/ESI MS, *J. Am. Soc. Mass Spectrom.* 14, 506–515.
21. Griffith, W. P., and Kaltashov, I. A. (2003) Highly asymmetric interactions between globin chains during hemoglobin assembly revealed by electrospray ionization mass spectrometry, *Biochemistry* 42, 10024–10033.
22. Griffith, W. P. and Kaltashov, I. A. (2007) Protein conformational heterogeneity as a binding catalyst: ESI-MS study of hemoglobin H formation, *Biochemistry* 46, 2020–2026.
23. Kamadurai, H. B., Subramaniam, S., Jones, R. B., Green-Church, K. B., and Foster, M. P. (2003) Protein folding coupled to DNA binding in the catalytic domain of bacteriophage lambda integrase detected by mass spectrometry, *Protein Sci.* 12, 620–626.
24. Guy, P., Remigy, H., Jaquinod, M., Bersch, B., Blanchard, L., Dolla, A., and Forest, E. (1996) Study of the new stability properties induced by amino acid replacement of tyrosine 64 in cytochrome C553 from *Desulfovibrio vulgaris* Hildenborough using electrospray ionization mass spectrometry, *Biochem. Biophys. Res. Commun.* 218, 97–103.
25. Snijder, J., Rose, R. J., Raijmakers, R., and Heck, A. J. R. (2011) Site-specific methionine oxidation in calmodulin affects structural integrity and interaction with Ca^{2+}/calmodulin-dependent protein kinase II, *J. Struct. Biol.* 174, 187–195.
26. Liu, J. and Konermann, L. (2009) Irreversible thermal denaturation of cytochrome *c* studied by electrospray mass spectrometry, *J. Am. Soc. Mass. Spectrom.* 20, 819–828.
27. Zhang, Y. H., Yan, X., Maier, C. S., Schimerlik, M. I., and Deinzer, M. L. (2001) Structural comparison of recombinant human macrophage colony stimulating factor beta and a partially reduced derivative using hydrogen deuterium exchange and electrospray ionization mass spectrometry, *Protein Sci.* 10, 2336–2345.
28. Zamani, L., Lindholm, J., Ilag, L. L., and Jacobsson, S. P. (2009) Discrimination among IgG1-κ monoclonal antibodies produced by two cell lines using charge state distributions in

nanoESI-TOF mass spectra, *J. Am. Soc. Mass Spectrom.* 20, 1030–1036.

29. Bobst, C. E., Thomas, J. J., Salinas, P., Savickas, P., and Kaltashov, I. A. (2010) Impact of oxidation on protein therapeutics: Conformational dynamics of intact and oxidized acid-β-glucocerebrosidase at near-physiological pH, *Protein Sci.* 19, 2366–2378.
30. Bobst, C. E., Abzalimov, R. R., Houde, D., Kloczewiak, M., Mhatre, R., Berkowitz, S. A., and Kaltashov, I. A. (2008) Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches, *Anal Chem* 80, 7473–7481.
31. Boys, B. L., Kuprowski, M. C., and Konermann, L. (2007) Symmetric behavior of hemoglobin alpha- and beta-subunits during acid-induced denaturation observed by electrospray mass spectrometry, *Biochemistry* 46, 10675–10684.
32. Kaltashov, I. A., Bobst, C. E., Abzalimov, R. R., Berkowitz, S. A., and Houde, D. (2010) Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry, *J. Am. Soc. Mass Spectrom.* 21, 323–337.
33. Kaltashov, I. A. and Abzalimov, R. R. (2008) Do ionic charges in ESI MS provide useful information on macromolecular structure?, *J. Am. Soc. Mass Spectrom.* 19, 1239–1246.
34. Dobo, A. and Kaltashov, I. A. (2001) Detection of multiple protein conformational ensembles in solution via deconvolution of charge state distributions in ESI MS, *Anal. Chem.* 73, 4763–4773.
35. Mohimen, A., Dobo, A., Hoerner, J. K., and Kaltashov, I. A. (2003) A chemometric approach to detection and characterization of multiple protein conformers in solution using electrospray ionization mass spectrometry, *Anal. Chem.* 75, 4139–4147.
36. Hendler, R. W. and Shrager, R. I. (1994) Deconvolutions based on singular value decomposition and the pseudoinverse: a guide for beginners, *J. Biochem. Biophys. Methods* 28, 1–33.
37. Frimpong, A. K., Abzalimov, R. R., Uversky, V. N. and Kaltashov, I. A. (2010) Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: Conformational heterogeneity of alpha-synuclein, *Proteins* 78, 714–722.
38. Testa, L., Brocca, S., Šamalikova, M., Santambrogio, C., Alberghina, L., and Grandori, R. (2011) Electrospray ionization-mass spectrometry conformational analysis of isolated domains of an intrinsically disordered protein, *Biotechnol. J.* 6, 96–100.
39. Brocca, S., Testa, L., Sobott, F., Samalikova, M., Natalello, A., Papaleo, E., Lotti, M., De Gioia, L., Doglia, Silvia M., Alberghina, L., and Grandori, R. (2011) Compaction properties of an intrinsically disordered protein: Sic1 and its kinase-inhibitor domain, *Biophys. J.* 100, 2243–2252.
40. Iavarone, A. T., Jurchen, J. C., and Williams, E. R. (2000) Effects of solvent on the maximum charge state and charge state distribution of protein ions produced by electrospray ionization, *J. Am. Soc. Mass Spectrom.* 11, 976–985.
41. Iavarone, A. T. and Williams, E. R. (2002) Supercharging in electrospray ionization: effects on signal and charge, *Int. J. Mass Spectrom.* 219, 63–72.
42. Schnier, P. D., Gross, D. S., and Williams, E. R. (1995) On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization, *J. Am. Soc. Mass Spectrom.* 6, 1086–1097.
43. Gumerov, D. R., Dobo, A., and Kaltashov, I. A. (2002) Protein-ion charge-state distributions in electrospray ionization mass spectrometry: distinguishing conformational contributions from masking effects, *Eur. J. Mass Spectrom.* 8, 123–129.
44. Frimpong, A. K., Abzalimov, R. R., Eyles, S. J., and Kaltashov, I. A. (2007) Gas-phase interference-free analysis of protein ion charge-state distributions: detection of small-scale conformational transitions accompanying pepsin inactivation, *Anal. Chem.* 79, 4154–4161.
45. Campos, L. A. and Sancho, J. (2003) The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH, *FEBS Lett*. 538, 89–95.
46. Konno, T., Kamatari, Y. O., Tanaka, N., Kamikubo, H., Dobson, C. M., and Nagayama, K. (2000) A partially unfolded structure of the alkaline-denatured state of pepsin and its implication for stability of the zymogen-derived protein, *Biochemistry* 39, 4182–4190.
47. Kaltashov, I. A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal. Chem.* 77, 5370–5379.
48. Testa, L., Brocca, S., and Grandori, R. (2011) Charge-surface correlation in electrospray ionization of folded and unfolded proteins, *Anal. Chem.* 83, 6459–6463.
49. Abzalimov, R. R., Frimpong, A. K., and Kaltashov, I. A. (2006) Gas-phase processes and measurements of macromolecular properties in solution: On the possibility of false positive and false negative signals of protein unfolding, *Int. J. Mass Spectrom.* 253, 207–216.
50. Jurchen, J. C., Garcia, D. E., and Williams, E. R. (2003) Gas-phase dissociation pathways of multiply charged peptide clusters, *J. Am. Soc. Mass Spectrom.* 14, 1373–1386.
51. Nemes, P., Goyal, S., and Vertes, A. (2008) Conformational and noncovalent complexation changes in proteins during electrospray ionization, *Anal. Chem.* 80, 387–395.
52. Kuprowski, M. C. and Konermann, L. (2007) Signal response of coexisting protein conformers in electrospray mass spectrometry, *Anal. Chem.* 79, 2499–2506.
53. Guevremont, R., Siu, K. W. M., Le Blanc, J. C. Y., and Berman, S. S. (1992) Are the electrospray mass spectra of proteins related to their aqueous solution chemistry?, *J. Am. Soc. Mass Spectrom.* 3, 216–224.
54. Leblanc, J. C. Y., Guevremont, R., and Siu, K. W. M. (1993) Electrospray mass-spectrometry of some proteins and the aqueous-solution acid-base-equilibrium model in the negative-ion detection mode, *Int. J. Mass Spectrom. Ion Proc.* 125, 145–153.
55. Kelly, M. A., Vestling, M. M., Fenselau, C. C., and Smith, P. B. (1992) Electrospray analysis of proteins: a comparison

of positive-ion and negative-ion mass spectra at high and low pH, *Org. Mass Spectrom.* 27, 1143–1147.

56. Peschke, M., Verkerk, U. H., and Kebarle, P. (2004) Prediction of the charge states of folded proteins in electrospray ionization, *Eur. J. Mass Spectrom.* 10, 993–1002.
57. Craig, P. O., Ureta, D. B., and Delfino, J. M. (2002) Probing protein conformation with a minimal photochemical reagent, *Protein Sci.* 11, 1353–1366.
58. Craig, P. O., Gomez, G. E., Ureta, D. B., Caramelo, J. J., and Delfino, J. M. (2009) Experimentally approaching the solvent-accessible surface area of a protein: Insights into the acid molten globule of bovine alpha-lactalbumin, *J. Mol. Biol.* 394, 982–993.
59. Stocks, B. B. and Konermann, L. (2009) Structural characterization of short-lived protein unfolding intermediates by laser-induced oxidative labeling and mass spectrometry, *Anal. Chem.* 81, 20–27.
60. Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Anal. Chem.* 81, 6563–6571.
61. Kiselar, J. G. and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting, *J. Mass Spectrom.* 45, 1373–1382.
62. Tong, X., Wren, J. C., and Konermann, L. (2008) γ-Ray-mediated oxidative labeling for detecting protein conformational changes by electrospray mass spectrometry, *Anal. Chem.* 80, 2222–2231.
63. Maleknia, S. D., Downard, K. M., and Chance, M. R. (2000) Synchrotron X-ray footprinting reveals unfolding of apomyoglobin helices, *Biophys. J.* 78, 44a–44a.
64. Chance, M. R. (2001) Unfolding of apomyoglobin examined by synchrotron footprinting, *Biochem. Biophys. Res. Comm.* 287, 614–621.
65. Narayan, M., Welker, E., Wedemeyer, W. J., and Scheraga, H. A. (2000) Oxidative folding of proteins, *Acc. Chem. Res.* 33, 805–812.
66. Narayan, M., Welker, E., and Scheraga, H. A. (2003) Characterizing the unstructured intermediates in oxidative folding *Biochemistry* 42, 6947–6955.
67. Creighton, T. E. (1997) Protein folding coupled to disulphide bond formation, *Biol. Chem.* 378, 731–744.
68. van den Berg, B., Chung, E. W., Robinson, C. V., and Dobson, C. M. (1999) Characterisation of the dominant oxidative folding intermediate of hen lysozyme, *J. Mol. Biol.* 290, 781–796.
69. Wu, J. and Watson, J. T. (1997) A novel methodology for assignment of disulfide bond pairings in proteins, *Protein Sci.* 6, 391–398.
70. Wu, J. and Watson, J. T. (1998) Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping, *Anal. Biochem.* 258, 268–276.
71. Yang, Y., Wu, J., and Watson, J. T. (1998) Disulfide mass mapping in proteins containing adjacent cysteines is possible with cyanylation/cleavage methodology, *J. Am. Chem. Soc.* 120, 5834–5835.
72. Wu, J., Yang, Y., and Watson, J. T. (1998) Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry, *Protein Sci.* 7, 1017–1028.
73. Yang, Y., Wu, J., and Watson, J. T. (1999) Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3) IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry, *J. Biol. Chem.* 274, 37598–37604.
74. Happersberger, H. P., Przybylski, M., and Glocker, M. O. (1998) Selective bridging of bis-cysteinyl residues by arsonous acid derivatives as an approach to the characterization of protein tertiary structures and folding pathways by mass spectrometry, *Anal. Biochem.* 264, 237–250.
75. Happersberger, H. P., Cowgill, C., and Glocker, M. O. (2002) Structural characterization of monomeric folding intermediates of recombinant human macrophage-colony stimulating factor β(rhM-CSF) by chemical trapping, chromatographic separation and mass spectrometric peptide mapping, *J. Chromatog. B-Analy. Technol. Biomed. Life Sci.* 782, 393–404.
76. Wang, Y., Lu, Q., Wu, S.-L., Karger, B. L., and Hancock, W. S. (2011) Characterization and comparison of disulfide linkages and scrambling patterns in therapeutic monoclonal antibodies: Using LC–MS with electron transfer dissociation, *Anal. Chem.* 83, 3133–3140.
77. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) Hydrogen exchange: the modern legacy of Linderstrøm-Lang, *Protein Sci.* 6, 1101–1109.
78. Englander, S. W. (2000) Protein folding intermediates and pathways studied by hydrogen exchange, *Annu. Rev. Biophys. Biomol. Struct.* 29, 213–238.
79. Englander, S. W., Mayne, L., and Rumbley, J. N. (2002) Submolecular cooperativity produces multi-state protein unfolding and refolding, *Biophys. Chem. 101–102*, 57–65.
80. Katta, V. and Chait, B. T. (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, *Rapid Commun. Mass Spectrom.* 5, 214–217.
81. Sobott, F. and Robinson, C. V. (2002) Protein complexes gain momentum, *Curr. Opin. Struct. Biol.* 12, 729–734.
82. Dempsey, C. E. (2001) Hydrogen exchange in peptides and proteins using NMR-spectroscopy, *Progr. Nucl. Magn. Res. Spectrosc.* 39, 135–170.
83. Connelly, G. P., Bai, Y. W., Jeng, M. F., and Englander, S. W. (1993) Isotope effects in peptide group hydrogen exchange, *Proteins* 17, 87–92.
84. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins* 17, 75–86.
85. Liepinsh, E. and Otting, G. (1996) Proton exchange rates from amino acid side chains - implications for image contrast, *Magn. Reson. Med.* 35, 30–42.
86. Hvidt, A. and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Adv. Protein Chem.* 21, 287–386.

87. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) Protein stability parameters measured by hydrogen exchange, *Proteins* 20, 4–14.

88. Maity, H., Lim, W. K., Rumbley, J. N., and Englander, S. W. (2003) Protein hydrogen exchange mechanism: local fluctuations, *Protein Sci.* 12, 153–160.

89. Qian, H., and Chan, S. I. (1999) Hydrogen exchange kinetics of proteins in denaturants: a generalized two-process model, *J. Mol. Biol.* 286, 607–616.

90. Itzhaki, L. S., Neira, J. L., and Fersht, A. R. (1997) Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature, *J. Mol. Biol.* 270, 89–98.

91. Milne, J. S., Mayne, L., Roder, H., Wand, A. J., and Englander, S. W. (1998) Determinants of protein hydrogen exchange studied in equine cytochrome c, *Protein Sci.* 7, 739–745.

92. Miller, D. W. and Dill, K. A. (1995) A statistical mechanical model for hydrogen exchange in globular proteins, *Protein Sci.* 4, 1860–1873.

93. Xiao, H., Hoerner, J. K., Eyles, S. J., Dobo, A., Voigtman, E., Mel'cuk, A. I., and Kaltashov, I. A. (2005) Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment, *Protein Sci.* 14, 543–557.

94. Kaltashov, I. A. (2005) Probing protein dynamics and function under native and mildly denaturing conditions with hydrogen exchange and mass spectrometry, *Int. J. Mass Spectrom.* 240, 249–259.

95. Hoerner, J. K., Xiao, H., and Kaltashov, I. A. (2005) Structural and dynamic characteristics of a partially folded state of ubiquitin revealed by hydrogen exchange mass spectrometry, *Biochemistry* 44, 11286–11294.

96. Eyles, S. J., Dresch, T., Gierasch, L. M., and Kaltashov, I. A. (1999) Unfolding dynamics of a β-sheet protein studied by mass spectrometry, *J. Mass Spectrom.* 34, 1289–1295.

97. Arrington, C. B. and Robertson, A. D. (2000) Correlated motions in native proteins from MS analysis of NH exchange: Evidence for a manifold of unfolding reactions in ovomucoid third domain, *J. Mol. Biol.* 300, 221–232.

98. Arrington, C. B., Teesch, L. M., and Robertson, A. D. (1999) Defining protein ensembles with native-state NH exchange: Kinetics of interconversion and cooperative units from combined NMR and MS analysis, *J. Mol. Biol.* 285, 1265–1275.

99. Rosa, J. J. and Richards, F. M. (1979) An experimental procedure for increasing the structural resolution of chemical hydrogen-exchange measurements on proteins: application to ribonuclease S peptide, *J. Mol. Biol.* 133, 399–416.

100. Englander, J. J., Rogero, J. R., and Englander, S. W. (1985) Protein hydrogen exchange studied by the fragment separation method, *Anal. Biochem.* 147, 234–244.

101. Zhang, Z. and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation, *Protein Sci.* 2, 522–531.

102. Wang, L., Pan, H., and Smith, D. L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions, *Mol. Cel. Proteomics* 1, 132–138.

103. Ehring, H. (1999) Hydrogen exchange/electrospray ionization mass spectrometry studies of structural features of proteins and protein/protein interactions, *Anal. Biochem.* 267, 252–259.

104. Brier, S., Maria, G., Carginale, V., Capasso, A., Wu, Y., Taylor, R. M., Borotto, N. B., Capasso, C., and Engen, J. R. (2007) Purification and characterization of pepsins A1 and A2 from the Antarctic rock cod Trematomus bernacchii, *FEBS J.* 274, 6152–6166.

105. Cravello, L., Lascoux, D., and Forest, E. (2003) Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins, *Rapid Commun. Mass Spectrom.* 17, 2387–2393.

106. Rey, M., Man, P., Brandolin, G., Forest, E., and Pelosi, L. (2009) Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry, *Rapid Commun. Mass Spectrom.* 23, 3431–3438.

107. Marcoux, J., Thierry, E., Vives, C., Signor, L., Fieschi, F., and Forest, E. (2010) Investigating alternative acidic proteases for H/D exchange coupled to mass spectrometry: Plasmepsin 2 but not plasmepsin 4 is active under quenching conditions, *J. Am. Soc. Mass. Spectrom.* 21, 76–79.

108. Zhang, H. M., Bou-Assaf, G. M., Emmett, M. R., and Marshall, A. G. (2008) Fast reversed-phase liquid chromatography to reduce back exchange and increase throughput in H/D exchange monitored by FT-ICR mass spectrometry, *J. Am. Soc. Mass Spectrom.* 20, 520–524.

109. Bai, Y. W., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen-exchange, *Proteins* 17, 75–86.

110. Woods, V. L., Jr. and Hamuro, Y. (2001) High resolution, high-throughput amide deuterium exchange-mass spectrometry (DXMS) determination of protein binding site structure and dynamics: utility in pharmaceutical design, *J. Cell. Biochem.* 37S, 89–98.

111. Burke, J. E., Karbarz, M. J., Deems, R. A., Li, S., Woods, V. L., and Dennis, E. A. (2008) Interaction of group IA phospholipase A2 with metal ions and phospholipid vesicles probed with deuterium exchange mass spectrometry, *Biochemistry* 47, 6451–6459.

112. Jones, L. M., Zhang, H., Vidavsky, I., and Gross, M. L. (2010) Online, high-pressure digestion system for protein characterization by hydrogen/deuterium exchange and mass spectrometry, *Anal. Chem.* 82, 1171–1174.

113. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L., Jr. (2005) Structure and properties of α-synuclein and other amyloids determined at the amino acid level, *Proc. Natl. Acad. Sci. USA.* 102, 15477–15482.

114. Deng, Y. Z., Pan, H., and Smith, D. L. (1999) Selective isotope labeling demonstrates that hydrogen exchange at

individual peptide amide linkages can be determined by collision-induced dissociation mass spectrometry, *J. Am. Chem. Soc.* 121, 1966–1967.

115. Kim, M. Y., Maier, C. S., Reed, D. J., and Deinzer, M. L. (2001) Site-specific amide hydrogen/deuterium exchange in *E. coli* thioredoxins measured by electrospray ionization mass spectrometry, *J. Am. Chem. Soc.* 123, 9860–9866.
116. Jørgensen, T. J., Gardsvoll, H., Ploug, M., and Roepstorff, P. (2005) Intramolecular migration of amide hydrogens in protonated peptides upon collisional activation, *J. Am. Chem. Soc.* 127, 2785–2793.
117. Rand, K. D., Adams, C. M., Zubarev, R. A., and Jørgensen, T. J. D. (2008) Electron capture dissociation proceeds with a low degree of intramolecular migration of peptide amide hydrogens, *J. Am. Chem. Soc.* 130, 1341–1349.
118. Zehl, M., Rand, K. D., Jensen, O. N., and Jørgensen, T. J. (2008) Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution, *J. Am. Chem. Soc.* 130, 17453–17459.
119. Rand, K. D., Zehl, M., Jensen, O. N., and Jørgensen, T. J. D. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry, *Anal. Chem. 81* 5577–5584.
120. Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) Measurement of amide hydrogen exchange by MALDI–TOF mass spectrometry, *Anal. Chem.* 70, 3987–3995.
121. Anderegg, R. J., Wagner, D. S., Stevenson, C. L., and Borchardt, R. (1994) The mass spectrometry of helical unfolding in peptides, *J. Am. Soc. Mass Spectrom.* 5, 425–433.
122. Waring, A. J., Mobley, P. W., and Gordon, L. M. (1998) Conformational mapping of a viral fusion peptide in structure-promoting solvents using circular dichroism and electrospray mass spectrometry, *Proteins* S2 38–49.
123. Kraus, M., Janek, K., Bienert, M., and Krause, E. (2000) Characterization of intermolecular beta-sheet peptides by mass spectrometry and hydrogen isotope exchange, *Rapid Commun. Mass Spectrom.* 14, 1094–1104.
124. Demmers, J. A., Haverkamp, J., Heck, A. J., Koeppe, R. E., 2nd, and Killian, J. A. (2000) Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers, *Proc. Natl. Acad. Sci. USA* 97, 3189–3194.
125. Demmers, J. A., van Duijn, E., Haverkamp, J., Greathouse, D. V., Koeppe, R. E., 2nd, Heck, A. J., and Killian, J. A. (2001) Interfacial positioning and stability of transmembrane peptides in lipid bilayers studied by combining hydrogen/deuterium exchange and mass spectrometry, *J. Biol. Chem.* 276, 34501–34508.
126. Akashi, S., Naito, Y., and Takio, K. (1999) Observation of hydrogen-deuterium exchange of ubiquitin by direct analysis of electrospray capillary-skimmer dissociation with Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* 71, 4974–4980.
127. Kaltashov, I. A., Bobst, C. E., and Abzalimov, R. R. (2009) H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: Is there a need for a top-down approach?, *Anal. Chem.* 81, 7892–7899.
128. Abzalimov, R. R. and Kaltashov, I. A. (2006) Extraction of local hydrogen exchange data from HDX CAD MS measurements by deconvolution of isotopic distributions of fragment ions, *J. Am. Soc. Mass Spectrom.* 17, 1543–1451.
129. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2011) Conformer-specific hydrogen exchange analysis of Aβ(1-42) oligomers by top-down electron capture dissociation mass spectrometry, *Anal. Chem.* 83, 5386–5393.
130. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2010) Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry, *Anal. Chem.* 82, 8591–8597.
131. Ferguson, P. L., Pan, J., Wilson, D. J., Dempsey, B., Lajoie, G., Shilton, B., and Konermann, L. (2007) Hydrogen/deuterium scrambling during quadrupole time-of-flight MS/MS analysis of a zinc-binding protein domain, *Anal. Chem.* 79, 153–160.
132. Hoerner, J. K., Xiao, H., Dobo, A., and Kaltashov, I. A. (2004) Is there hydrogen scrambling in the gas phase? Energetic and structural determinants of proton mobility within protein ions, *J. Am. Chem. Soc.* 126, 7709–7717.
133. Eyles, S. J., Speir, J. P., Kruppa, G. H., Gierasch, L. M., and Kaltashov, I. A. (2000) Protein conformational stability probed by Fourier transform ion cyclotron resonance mass spectrometry, *J. Am. Chem. Soc.* 122, 495–500.
134. Abzalimov, R. R. and Kaltashov, I. A. (2010) Controlling hydrogen scrambling in multiply charged protein ions during collisional activation: Implications for top-down hydrogen/deuterium exchange MS utilizing collisional activation in the gas phase, *Anal. Chem.* 82, 942–950.
135. Kaltashov, I. A. and Eyles, S. J. (2002) Crossing the phase boundary to study protein dynamics and function: combination of amide hydrogen exchange in solution and ion fragmentation in the gas phase, *J. Mass Spectrom.* 37, 557–565.
136. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2008) Electron capture dissociation of electrosprayed protein ions for spatially resolved hydrogen exchange measurements, *J. Am. Chem. Soc.* 130, 11574–11575.
137. Abzalimov, R. R., Kaplan, D. A., Easterling, M. L., and Kaltashov, I. A. (2009) Protein conformations can be probed in top-down HDX MS experiments utilizing electron transfer dissociation of protein ions without hydrogen scrambling, *J. Am. Soc. Mass Spectrom.* 20, 1514–1517.
138. Zubarev, R. A., Zubarev, A. R., and Savitski, M. M. (2008) Electron capture/transfer versus collisionally activated/induced dissociations: solo or duet?, *J. Am. Soc. Mass Spectrom.* 19, 753–761.
139. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2009) Hydrogen/deuterium exchange mass spectrometry with top-down electron capture dissociation for characterizing

structural transitions of a 17 kDa protein, *J. Am. Chem. Soc.* 131, 12801–12808.

140. Kaltashov, I. A., Eyles, S. J., and Xiao, H. (2004) Combination of protein hydrogen exchange and tandem mass spectrometry as an emerging tool to probe protein structure, dynamics and function, In *Focus on Protein Research* (Robinson, J. W. Ed.), pp 191–218, Nova Science Publishers, Inc., Hauppauge, NY.
141. Marshall, A. G., Senko, M. W., Li, W., Li, M., Dillon, S., Guan, S., and Logan, T. M. (1997) Protein molecular mass to 1 Da by ^{13}C, ^{15}N double-depletion and FT ICR mass spectrometry., *J. Am. Chem. Soc.* 119, 433–434.
142. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through and obligatory intermediate, *Protein Sci.* 8, 45–49.
143. Konermann, L. and Simmons, D. A. (2003) Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry, *Mass Spectrom. Rev.* 22, 1–26.

6

KINETIC STUDIES BY MASS SPECTROMETRY

In Chapter 5, we demonstrated the power of mass spectrometry (MS) for the study of biomolecules at equilibrium. We can allow a sample to equilibrate in the solution conditions of choice and then acquire a spectrum at leisure, which will reveal the presence of non-native states and/or provide their thermodynamic or structural characteristics. Intermediate states can also be populated transiently in a variety of processes ranging from protein folding to enzyme catalysis. However, their detection and characterization, the prime targets of kinetic studies, are usually much more challenging due to their fleeting nature. Kinetic studies are crucial to a complete understanding of biochemical reaction mechanisms, including protein folding, protein assembly, and enzyme function. Kinetic investigations enable rate constants and activation energies to be derived, and by monitoring the progress of the reaction it may be possible to observe transient intermediates that give key mechanistic information. Transient intermediates detected during protein folding can provide insight into the dominant pathway of folding; likewise short-lived intermediate species detected during the catalytic cycle of enzyme activity give valuable information about the reaction mechanism. However, these intermediates are, by their very nature, short-lived so, unlike equilibrium measurements, there is the added challenge of time scale. Once a reaction is initiated it must be rapidly transferred to an observation point in order to be followed kinetically. In this chapter, we will discuss some of the methodologies developed to study these kinetic events.

6.1. KINETICS OF PROTEIN FOLDING

Perhaps the simplest way to measure the kinetics of a reaction is to mix the reactants and then monitor change with time of a spectroscopic signal. In the case of protein folding, this involves removing chemical denaturant by dilution into conditions permissive for refolding, or adjustment of pH, for instance. Proteins fold to their native state rapidly under refolding conditions, generally on the order of only a few seconds. Thus a method is needed that can monitor this reaction on a relatively short time scale. Instrumentation for *continuous-flow* and *stopped-flow* measurement of enzyme kinetics was originally developed in the 1930s (1,2).

6.1.1. Stopped-Flow Measurement of Kinetics

The basic stopped-flow experiment involves rapid mixing of reactants in a controlled ratio with subsequent placement of the reaction mixture in a flow cell, allowing the reaction to be monitored by optical methods. In the simplest folding experiment, a syringe is loaded with protein denatured chemically by a high concentration of chaotrope. A second syringe contains buffer and the two are connected to a mixing device (Fig. 6.1a). Upon activation of the two syringes, denatured protein is mixed with refolding buffer, diluting the denaturant such that the protein begins to fold. The reaction mixture then flows through to an observation cell and the flow is stopped, such that the progress of the reaction can be monitored as a function of time by a variety

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

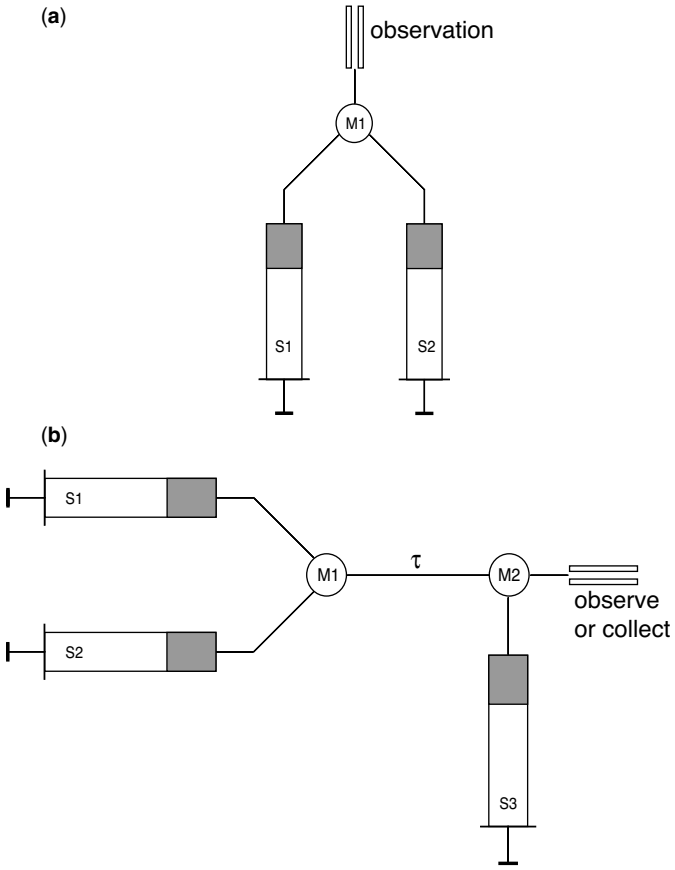


Figure 6.1. (*a*) Schematic diagram of a stopped flow device. Reactants contained in the two drive syringes (S1 and S2) are mixed in mixer M1 when the syringes are activated, and the reaction mixture is then transferred to an observation cell where the reaction kinetics can be monitored in real-time spectroscopically. (*b*) In a quench flow device the reaction mixture emerging from M1 is allowed to react for a variable time period, τ, before quenching by mixing at M2 with either a quenching agent or by injecting into liquid nitrogen.

of spectroscopic methods (see Chapter 2), each of which monitors different properties of the folding protein. Commonly, circular dichroism (CD) is employed to track secondary [in the far-ultraviolet (UV) wavelength region] or tertiary (near-UV) structure formation; fluorescence is employed to measure the environment of tryptophan residues or extrinsic fluorophores; optical absorption measurements can be used at a wavelength where there is a change in extinction coefficient upon folding.

Commercial stopped-flow instrumentation that ensures accurate mixing ratios by employing syringes controlled by stepping motors or pneumatic drive systems has been available for some time. Dead times are minimized by use of high-efficiency mixers that provide turbulent flow and minimal flow paths between mixer and observation cell. Additionally, stop syringes or valves provide a precise stop of the flow within the cell so that turbulence effects during observation are minimized. Typically, dead times in the millisecond range are readily achievable, enabling folding reactions that typically have time scales from several milliseconds to several seconds to be monitored. Even more rapid folding kinetics can be measured using high efficiency mixers coupled to continuous-flow measurement devices (3), as described in Section 6.2.2.

Stopped-flow measurements by optical spectroscopy require the presence of a chromophore whose properties change as the reaction occurs (see Chapter 2). An alternative and, in some cases, much more tractable technique involves quenching the reaction at some point during its progress, followed by offline analysis of the species present in the quenched mixture. Analysis can then be carried out using a variety of different methods that are not dependent on a chromophore. This effectively allows a snapshot to be taken of the species present at that precise time. While this does not allow kinetics to be measured directly, a kinetic curve can be produced by determining the species present in the quenched reaction at different time points. A *quenched-flow apparatus* involves not one but two mixing chambers linked by a delay loop (Fig. 6.1b). Reactants are introduced into the first mixing chamber (M1) and the reaction is allowed to occur in the delay line for a variable time period, followed by quenching of the reaction by mixing in the second chamber (M2). Alternatively, the reaction can be quenched by injecting the reaction mixture into liquid nitrogen. The species present at that time point can then be measured at leisure under conditions where no further reaction occurs. By varying the flow rate and/or volume of the delay loop, the reaction time allowed before quenching can be adjusted, enabling a kinetic time course to be reconstructed from individual reaction time points.

This type of methodology (normally employed over a longer time scale) has been employed for many decades to study enzymatic reactions by using, for instance, radio-labeled reagents in the initial mix followed by quenching with an excess of the unlabeled analogue. An example is the classic *pulse chase* experiment, which can be used for a variety of biochemical studies from localization of proteins within cells to investigation of protein translation on the ribosome, or to determine how metabolites are processed in response to different external factors (4–6). For example, protein synthesis under a certain cellular condition can be monitored by allowing cells to incubate for varying periods of time in the presence of ^{35}S-methionine, incorporating this radioisotope into newly translated peptide chains. After the "pulse" period the labeled amino acid is rinsed out and replaced by a "chase" buffer containing an excess of unlabeled methionine, effectively preventing further incorporation of radio-label into proteins translated beyond that point. Subsequent separation of the cellular components allows the identification and localization of those proteins produced during the labeling pulse by detection of radioactivity.

6.1.2. Kinetic Measurements with Hydrogen Exchange

The methods of isotope exchange at labile sites on proteins have been used for many years to investigate protein structure. From the previous chapters, we saw that labile protons (particularly backbone amides) are useful probes of structure and dynamics, since those that are buried or involved in hydrogen-bonding exchange more slowly with solvent than those that are exposed. Hydrogen–deuterium exchange (HDX) labeling has proven extremely powerful for studying the kinetics of protein folding (7,8). As an unfolded polypeptide forms structure during the folding reaction, labile amide protons become protected against exchange with bulk solvent due to burial or formation of intramolecular hydrogen bonds (i.e., as part of secondary structure). In a *competition experiment,* unfolded protein with all amides protiated is allowed to refold by rapid dilution into deuterated buffer. During this time, amides that are rapidly protected against exchange will remain protiated, whereas those that remain exposed to solvent will exchange with the buffer deuterons. After a certain time period, the reaction may be quenched by dropping the pH to a level where the intrinsic exchange rate is minimal (pH 2.5–3) and by reducing the temperature to 0°C. Refolding then continues under quenched conditions, where no further labeling can occur, thus enabling the regions of structure that become protected to be probed. In practice, it is a simpler matter to adjust the pH for the labeling reaction and allow for competition between folding and exchange rather than controlling the time of exposure to the labeling buffer. Since the intrinsic exchange rate changes, as a rule of thumb, by an order of magnitude for every pH unit (see Fig. 5.8), varying the pH of the labeling buffer during the folding reaction allows one to probe a variety of folding rates. As already discussed in the preceding chapters, the exchange process of labile protons with solvent is catalyzed by acid, base, and water, with the overall intrinsic rate given by $k_{\text{int}} = k_{\text{a}}[\text{H}^+] + k_{\text{b}}[\text{OH}^-] + k_{\text{w}}[\text{H}_2\text{O}]$, so that the intrinsic rate k_{int} can be calculated at a given pH (e.g., $t_{1/2} = 1/k_{\text{int}} \sim 1$ ms at pH 9.5, 25°C). In this manner, no special equipment is required, since the intrinsic rate of exchange can be matched to the folding rate, provided that the native structure is protected against exchange and folded regions in intermediate states are sufficiently stable at the chosen pH such that no further exchange occurs from these states. The folding reaction can be allowed to go more or less to completion before quenching and subsequent sample work-up for analysis.

To add a further level of complexity and even more power to probe folding pathways and kinetics, the *pulse-labeling experiment* employs a short labeling pulse after a variable refolding reaction time, followed by a third mixing step to quench exchange. This prevents further labeling, so that even marginally stable intermediates remain protected,

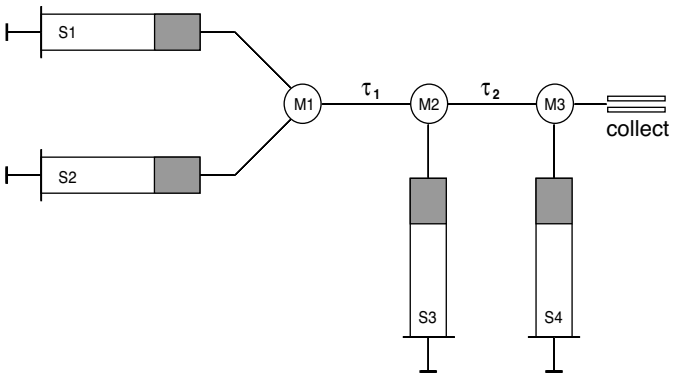


Figure 6.2. In the pulse-labeling refolding experiment, reactants from S1 and S2 are mixed to initiate the folding reaction for time τ_1. The labeling reagent from S3 is injected and reacted for time τ_2 before quenching in M3. The quenched reaction mixture is then collected for offline measurement.

and allows folding to continue to completion (Fig. 6.2). This method enables labeling to occur at given points during the reaction rather than continuously from initiation. The setup has become extremely popular as a tool for measuring protein-folding kinetics by hydrogen–deuterium labeling, taking advantage of the fact that exchange of labile hydrogens within proteins is extremely sensitive to pH. A high pH pulse will rapidly exchange all exposed amides, while if the labeling pulse is kept short, it generally has no significant effect on folding kinetics. At low pH, the intrinsic exchange rate is minimal, such that over the course of a folding reaction no significant further exchange with bulk solvent occurs at amide hydrogens on the time scale of minutes and tens of minutes, whereas in mildly basic solutions (pH > 9.5) the half-time for exchange is on the order of 1 ms.

Denatured protiated protein is first diluted by mixing with renaturation buffer at a pH where folding can occur. Refolding is allowed to proceed for a time τ_1, governed by the first delay loop (generally this delay can be varied from a few ms up to several s), followed by a short isotope-labeling pulse (τ_2) with a deuterated buffer at high pH. This pulse is used to label any amides with deuterium that are not sequestered from the solvent by burial or hydrogen bonding in the secondary structure. A 10 ms labeling pulse at pH 9.5 is equivalent to 10 half-lives of exchange so it is sufficient to completely label all exposed labile sites. The length of time for the labeling pulse is controlled by the second delay loop. This is followed by quenching in the third mixing chamber of any further HDX by reducing the pH to a level close to the minimum for intrinsic exchange (pH 2.5–3). By varying τ_1, a kinetic picture can be built up of the folding process based on the regions of structure that become protected against exchange.

The most common modern method to monitor deuterium incorporation from the pulse-labeling experiment described above is nuclear magnetic resonance (NMR), from which

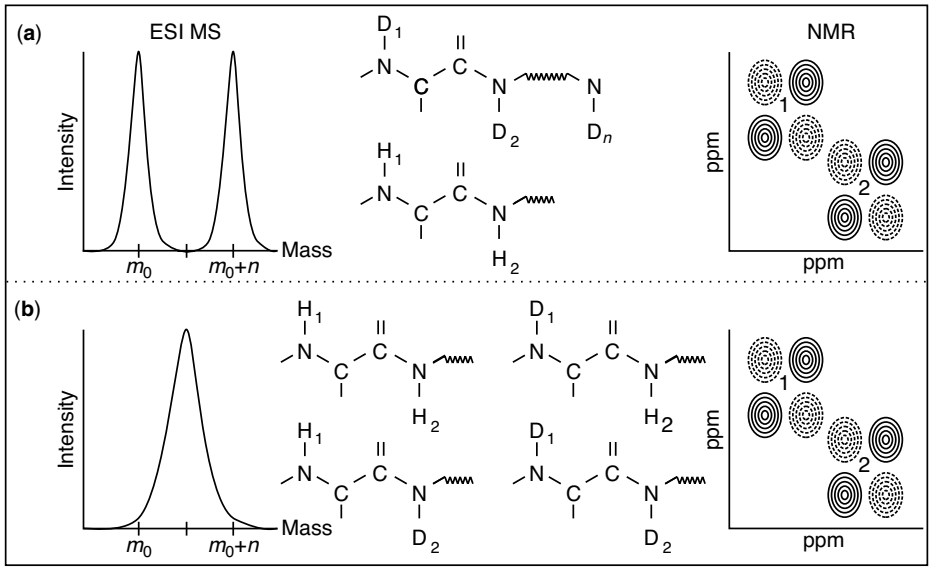


Figure 6.3. Distinguishing exchanging populations by MS. Simulated mass spectra and NMR spectra for a population of n amides with average 50% H/D exchange. (*a*) In 50% of the sample population, exchange for deuterium is complete, whereas for the remaining 50% no exchange has occurred (EX1 mechanism). The mass spectrum shows two peaks representing exchanged and unexchanged, and NMR shows 50% exchange at all amide sites. (*b*) 50% random exchange at each amide site (EX2 mechanism) produces an identical NMR spectrum, whereas MS now exhibits a single broad peak a average mass representing 50% H/D exchange. [Reproduced with permission from (11). Copyright © 1993 American Association for the Advancement of Science.]

the average proton occupancy at each amino acid can be measured as a function of time τ_1 during the folding process. By measuring peak intensities, a kinetic curve can be constructed for each residue, and the structural details of folding intermediates can be inferred from correlation of amides with similar protection rates. Over the past decades, this method has been applied extensively to mechanistic studies of the folding of a number of small proteins (8–10). The technique is limited, however, by the requirement for assignment of the NMR resonances in order to obtain meaningful structural information (see Section 2.3).

Another significant limitation arises from the fact that each NMR signal from a native protein sample corresponds to the sum of resonances from each molecule in the sample. For instance, the peak corresponding to the amide at position x in a protein sequence is the summation of resonance intensities of every amide at position x in every protein molecule in the NMR tube. In the case of a 0.5 mL sample containing 1 mM protein, this corresponds to $\sim 3 \times 10^{17}$ molecules! Thus, for an HDX experiment the measured NMR resonance intensity corresponds only to the average proton occupancy within the sample. As pointed out by Miranker et al. (11), NMR cannot distinguish between the different populations of amide protection present in a solution of native protein (Fig. 6.3). For instance, a sample containing molecules in which 50% are fully protonated and the remaining 50% are fully deuterated cannot be distinguished from a sample in which all the molecules in the sample are 50% randomly deuterated. In both cases, the NMR peak for a given amide would simply appear to have one-half of the intensity as that of the control protonated sample. This problem can be elegantly solved by using MS as the detection device. As is the case for coexisting equilibrium states (see Chapter 5), the differences in proton occupancy are easily resolved by their differences in molecular mass. In the former case, two peaks would be observable in the mass spectrum, one at the fully protonated mass (m_0), the other at the fully deuterated ($m_0 + n$). In the case of 50% random deuteration, a single peak would be observed centered at $m_0 + 0.5n$, as illustrated in Figure 6.3.

6.2. KINETICS BY MASS SPECTROMETRY

6.2.1. Pulse Labeling Mass Spectrometry

Since the pioneering work of Katta and Chait (12) demonstrated that differences in deuterium content could be measured in intact proteins by MS, there has been an upsurge in the use of electrospray ionization (ESI) MS in the study of protein-folding kinetics. The beauty of the technique and what makes it complementary to the information obtained by NMR methods is that individual populations with differing deuterium contents can be distinguished using MS by virtue

of the differences in molecular weight. This was first demonstrated by Miranker et al. (11) for the protein hen egg white lysozyme. Prior studies of this small protein using kinetic methods showed that a transient intermediate state populates during refolding under the conditions employed (10). Pulse labeling experiments revealed that residues in the α-helical domain become protected more rapidly during folding than in the β-sheet domain, but also, more interestingly, the kinetic traces for all of the amides monitored by NMR exhibited biphasic kinetics, with a fast phase rate constant similar for all probed amides. This suggested that there might be an alternative pathway of folding that rapidly protects the entire structure against exchange, effectively bypassing the intermediate. By adjusting the intensity (labeling time or pH) of the labeling pulse, the authors inferred the presence of parallel folding pathways, but no direct evidence could be obtained by NMR alone due to the aforementioned sample averaging of cross-peak intensities.

Mass spectrometry, on the other hand, can clearly distinguish these parallel folding events. If a population of folding molecules attains native-like structure, and hence complete protection against exchange, then it will have a different molecular weight from the unfolded state that becomes fully exchanged during the labeling pulse, or indeed any partially folded intermediate states. In the case of lysozyme refolding in a pulse labeling experiment, deuterated unfolded protein was allowed to refold for varying times before labeling with a ^1H_2O pulse. As shown in Fig. 6.4, even after only 20 ms of folding three peaks were observable in the mass spectrum corresponding to different degrees of exchange in the folding population. The low mass peak represents protein that remains unfolded and thus all labile sites were exchanged with protons. The two other peaks correspond to populations that have 22 and 50 amides protected, the former representing the folding intermediate while the latter arises from a parallel folding population in which complete protection is afforded to the entire molecule. At longer refolding times, as folding continues, the populations of the unfolded and intermediate states diminish as they are converted to native protected protein.

These elegant experiments demonstrate the power of MS to distinguish populations based on isotope incorporation during refolding. While unable to give high-resolution information as to the structural regions of the protein that become protected first, MS is unique in its ability to identify parallel folding pathways and other details about populations in an HDX sample that would be unattainable by NMR. Another huge advantage of ESI MS over NMR is that very little protein is required (routinely $< 1\%$ of that needed for NMR). In fact, the low concentration requirements ($< 10\,\mu$M) for ESI MS would, in principle, allow pulse-labeled samples to be directly infused into the mass spectrometer without further concentration provided that the buffers used to effect the pH changes are compatible.

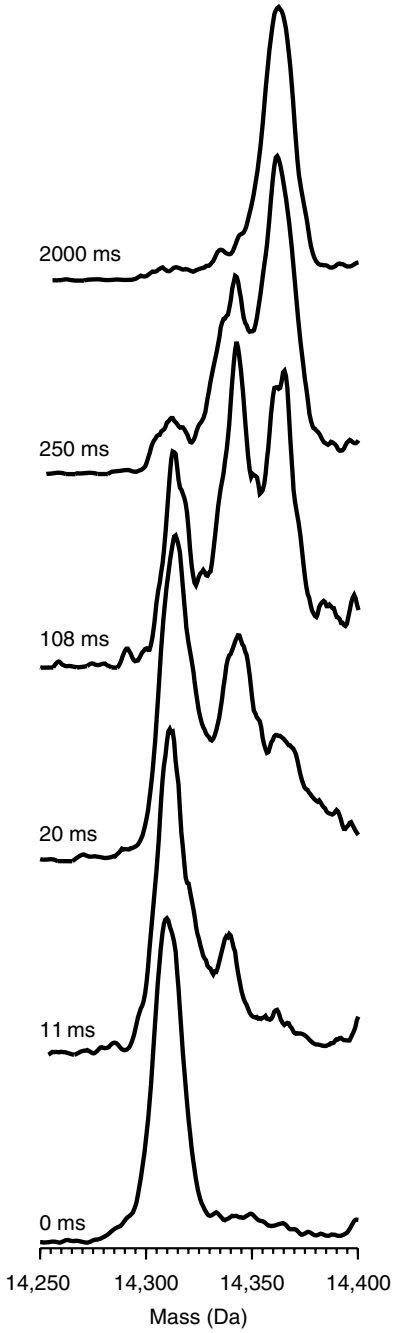


Figure 6.4. Time course for refolding of hen lysozyme measured by pulse labeling HDX MS. At the earliest labeling time point (0 ms), no protection exists so all protein molecules are fully labeled with ^1H. By contrast, after long refolding time (2000 ms) the protein becomes fully protected before the labeling pulse, so it remains fully deuterated. At intermediate time points, several conformational populations coexist, including fully protected and fully unprotected states, and a conformation with intermediate protection (degree of exchange). [Adapted with permission from (11). Copyright © 1993 American Association for the Advancement of Science.]

Alternatively, rapid column desalting could be used, in any case drastically reducing the time for workup and measurement, and thus minimizing the possible problems of back-exchange.

An interesting aspect of pulse labeling in combination with MS is the ability to distinguish not only different folding populations, but also altered kinetics in different proteins, if they are sufficiently different in mass that there is no overlap of peaks. Site-directed mutagenesis is often used to probe the influence of single amino acid residues on the folding and stability of proteins, but it can often be quite difficult to ensure absolutely identical refolding conditions are employed for each individual protein. Hooke et al. (13) noted that hen and human lysozyme, which have highly homologous amino acid sequences, but apparently quite different folding kinetics, could be refolded together as a mixture and then their HDX behavior monitored in the same experiment. This technique avoids the potential ambiguities that could arise from slightly different denaturation concentrations or mixing ratios during the refolding experiments, and allows very small changes in folding behavior to be assigned unequivocally. Of course, this option is not available to standard stopped-flow optical techniques unless extrinsic chromophores are added to the protein in order to distinguish them spectroscopically.

More recent studies have employed the HDX labeling strategy in conjunction with ESI MS to investigate the controversies of parallel pathways during protein folding. In many cases, experimental folding studies have clearly demonstrated the existence of intermediate species during the renaturation reactions for many proteins (14–18). For many years, it was assumed that intermediates might be essential precursors to the native structure. Characterization of these states has thus for some time been considered a key aspect for the better understanding of the processes by which a protein folds. Theoretical treatments and the development of the protein folding process in terms of the "new view" or landscape model have since, however, brought into question the relevance of these transient states. It has been suggested that these partially folded states may in fact merely be nonproductive kinetic traps that are off-pathway and not important to the folding process to the native state (19,20). A number of small proteins have been identified that appear to fold via a pure two-state mechanism involving no observable intermediates, suggesting that folding is a much more concerted process (21–25). In fact, experimental data for proteins that previously had been shown to have well-characterized intermediates have since suggested that proteins such as lysozyme or cytochrome *c* may in fact fold without kinetically detected intermediates under certain conditions (26,27).

The protein interleukin-1β is a small all β-sheet protein that folds very slowly (9), making it readily amenable to hydrogen-exchange studies. The protein refolds via a well-characterized intermediate based on HDX NMR and stopped-flow optical methods. However the question of whether this intermediate is an essential step during the folding of this protein was answered recently using MS (28). Pulse labeling HDX experiments measured by ESI MS demonstrated that there was no evidence for formation of any folded structure with native state protection at early time points. These experiments clearly showed that, for this protein at least, the native state does not become directly populated. Instead, there is a lag phase during which an intermediate species forms before formation of the native folded structure can occur. These mass spectrometric data suggest that, even in the landscape view of protein folding, there may be a preferred pathway of folding. Although the unfolded state samples a large amount of conformational space, there are only a limited number of ways to get to the native structure, and the intermediate structure detected is obligatory for correct folding.

A similar study was performed to investigate the folding of another small protein, apomyoglobin. This helical protein has been the subject of many folding studies by a variety of biophysical methods. The general consensus is that an intermediate state forms during folding, in which three of the native α-helices (the N-terminal A helix, and the two C-terminal G and H helices) fold and dock together prior to formation of the native structure. Once again, pulse labeling HDX experiments coupled to ESI MS analysis showed that no alternative route to the native state was observed (29). In fact, for this protein, within the first measured time point (6 ms), no unfolded protein was detected, but it had all been converted to an intermediate. Only at later time points was a mass spectral signal corresponding to fully protected native protein detectable. It has been suggested that partially folded states may not be on the preferred folding pathway, but are instead kinetic traps (27). If this were the case, one would expect to see formation of the native state in parallel with any nonproductive intermediates. In both of the above studies, this is not in fact the case, providing some evidence that, at least for these examples, intermediates are necessary and on the pathway of productive folding. In fact, for the globin family of proteins, despite diverse sequence differences, this folding mechanism may be conserved, as observed by Nishimura et al. (30).

In the case of lysozyme and interleukin-1β described above, the deuterium-labeled masses of the intermediate states are sufficiently different from those of either the native or unfolded states that they can be readily resolved in the mass spectra. In the experiments performed on myoglobin, however, the rapidly formed intermediate could not be completely resolved (29,31). In this case, the data were fit to a sum of Gaussian curves representing contributions from each of the three states (native, N, intermediate,

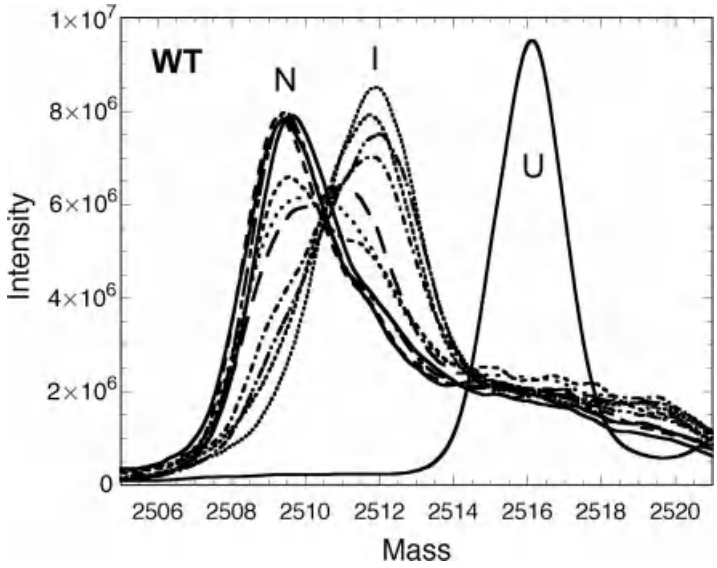


Figure 6.5. The HDX MS traces for refolding of apomyoglobin. No early peak arises corresponding to native-like protection suggesting an obligate intermediate species. In the case of apomyoglobin, the intermediate state is fully populated within the dead time of measurement (6 ms). The *m/z* values for unfolded, intermediate, and native state protections are very close and not fully resolved, but they can nevertheless be deconvoluted. [Reproduced with permission from (31). Copyright © 2003 Elsevier Ltd.]

I, and unfolded, U), based on complementary data from NMR experiments (32–34). It was found that satisfactory fits could be achieved for data at all time points, from which the time course of evolution of the intermediate and native states could be obtained (Fig. 6.5). Interestingly, the centroid masses and widths of the Gaussians were kept constant to fit the data at each time point. This result suggests a remarkable homogeneity and stability in protection of the intermediate state throughout the time course of the folding reaction, since no further changes in the exchange profile of this species were detectable, at least by the fitting procedure employed. More recent studies by NMR have sought to further investigate structure in the kinetic intermediate, and suggest a more heterogeneous ensemble that may be masked by back-exchange during the labeling pulse (35).

As mentioned above, one of the assumptions of the pulse labeling technique is that the folding pathway is not altered by the high pH labeling pulse. Marginally stable regions may well be destabilized to a sufficient extent that they too become labeled, and are hence not detected. Alternatively, if the folding rate is altered by a pH change, then this may give rise to incorrect apparent protection rates due to incomplete labeling (36). Thus, a great deal of care must be exercised in the interpretation of pulse labeling experiments. However, if adequate controls are performed and all factors are taken into consideration, the effects of varying pulse length and pulse pH can be valuably applied to investigate the conformational stability of partially folded intermediate species (8,37,38). A potential solution to the inherent difficulties has been provided by using continuous flow MS ((39), and Section 6.2.2).

A further enhancement of the pulse labeling method, namely subsequent fragmentation of the labeled protein prior to MS, enables more specific information to be obtained about the regions of the polypeptide chain protected during folding. The techniques of Yang and Smith (40) described in Chapter 5 may be readily applied to pulse labeled samples. After refolding, labeling, and quenching, the reaction mixture is subjected to peptic digest and the resulting proteolytic fragments are analyzed by high performance liquid chromatography (HPLC) coupled to ESI MS (Figure 6.6). This method was applied to samples of cytochrome *c* labeled after varying refolding times (40). The authors found that each peptic fragment was either completely labeled or unlabeled, indicating that folding in each segment was a highly cooperative process. The N- and C- terminal regions, however, were formed into a stable structure protective against exchange at earlier times than the remainder of the molecule (Fig. 6.7). These results were in accord with previous studies by NMR, but again MS provided valuable information about structural heterogeneity in the folding intermediates from the measured isotope pattern. In addition, this method could be extended to proteins of higher molecular weight, beyond the current capabilities of NMR methods.

An example of pulse labeling applied to the study of larger proteins is the case of rabbit muscle aldolase.

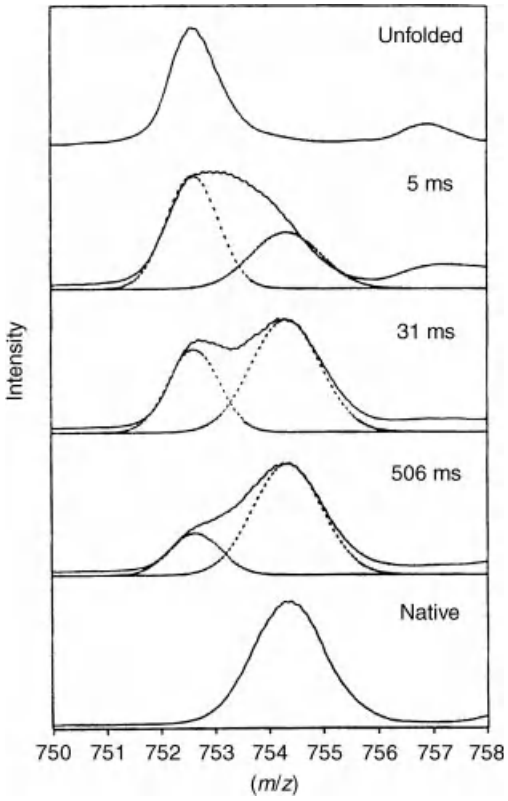


Figure 6.6. Proteolytic digestion after pulse labeling enables hydrogen-exchange protection information to be gleaned from the separate fragments. For cytochrome *c*, the N-terminal fragment, residues [1–21], becomes cooperatively protected during folding, with 50% of the population completely protected in this region within 30 ms. [Adapted with permission from (40). Copyright © 1997 American Chemical Society.]

This α-/β- barrel protein forms a homotetramer of molecular mass 157 kDa, far beyond the current capabilities of NMR investigation. Pan and Smith (41) employed a dialysis method to initiate refolding from the guanidine hydrochloride denatured state of this protein followed by pulse labeling of aliquots to study the refolding and reconstitution of this protein as a function of time. With the use of peptic digest to identify structural regions protected against exchange, the authors were able to delineate three structural domains, each consisting of $\sim$ 110 residues within each monomer, that folded sequentially. This finding is consistent with the current notion that larger proteins may fold initially to form subdomains of approximately this size (42,43), although the folding domains in this protein consist of noncontiguous segments. Interestingly, the authors observed that the domains differed in the folding and unfolding directions, suggesting that different intermediates might be more favored, although they attributed this primarily to stabilizing quaternary interactions in the folded state that may locally stabilize structural regions only in the unfolding pathway.

In principle, the top-down approaches discussed in Section 5.4.2 should also be applicable to the structural characterization of pulse labeled samples. Measurement of deuterium content in the intact protein by MS could be followed by subsequent localization to specific regions using the gas-phase fragmentation methods previously described. An important aspect of HDX studied by MS is that the protein does not have to be in its native state for measurement of deuteration by MS. For the NMR experiments, generally only the native state resonance assignments are available, so the final folded state must be formed prior to subsequent analysis, and under conditions for which NMR assignments have been made (NMR chemical shifts are highly dependent on pH and temperature). On the other hand, provided that the intrinsic exchange rate is minimal, deuterium labeling can be measured by MS under any conditions. Thus even though many proteins are only marginally stable at pH 2–3 they can be readily subjected to mass spectrometric analysis. These conditions may in fact be more favorable for promoting efficient gas-phase fragmentation, as a molecule with a higher charge state will have a greater collision energy, giving rise to lower fragmentation energy barriers.

There is a growing body of evidence that electron-based top-down fragmentation methods (e.g., ETD and ECD, described in Section 3.4.4) can be applied to measure

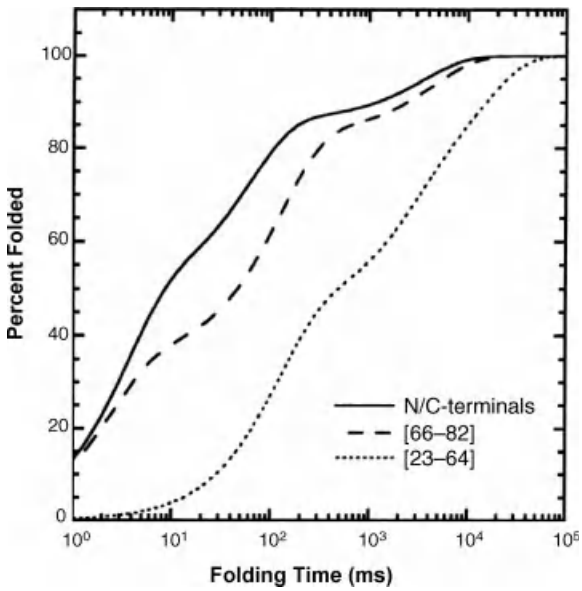


Figure 6.7. Time course of protection for various segments of cytochrome *c* during folding. The N- and C- termini become rapidly protected in the intermediate state encompassing the A, G, and H helices. In contrast the [23–64] segment folds significantly more slowly and the [66–82] region with kinetics between the two. [Reproduced with permission from (40). Copyright © 1997 American Chemical Society.]

deuterium content in a site-specific fashion without scrambling (44–46). The application of ECD to characterize intermediates in the folding of myoglobin was recently demonstrated by Pan et al. (47), who confirmed that the protection seen early in folding was indeed localized primarily to the G and H (and to a lesser extent A) helices. This result also opens the way to the possibility of selecting specific populations with different extents of deuterium incorporation, using, for instance, the front-end quadrupole of a hybrid Fourier transform ion cyclotron resonance (FT ICR) instrument, and probing the level of exchange in a site-specific fashion using gas-phase fragmentation techniques (48). This would enable the characteristics of individual intermediate ensembles to be probed in much greater detail than mass difference alone, opening the way to an even more complete description of structural heterogeneity in these states.

6.2.2. Continuous-Flow Mass Spectrometry

The beauty and relative simplicity of the pulse labeling method is that all of the folding and labeling is performed offline before introduction of the sample into the mass spectrometer. Thus, in the absence of back-exchange processes that can be relatively easily minimized by maintaining samples frozen until used, samples can be analyzed at leisure. However, even under carefully controlled conditions some information about residues protected during folding can be lost in the concentration and desalting processes due to exchange with solvent. For this reason, alternative methods have been sought that enable direct interface to the ESI source to enable monitoring of exchange online. In principle, this could not only improve the kinetic resolution, but also allow more amide probes to be monitored, even those that are only marginally stable under quench conditions.

The technique of *time-resolved* ESI MS was initially developed by Henion and co-workers (49), who used manual mixing followed by infusion of the reaction mixture into the spectrometer, and Sam et al. (50,51), to study reactions of small organic molecules. The latter setup is, in many ways, similar to the stopped-flow device employed for spectroscopic kinetic studies, except that the observation cell is replaced by a connection to the ESI interface. Since standard electrospray requires a continuous flow of analyte, it is not possible to stop the flow of liquid for observation (although see Section 6.2.3 for techniques in which this effectively can be achieved). Instead, the syringes are pushed continuously and the two reactants mix and then pass into a delay line. The total flow rate from the two syringes and the volume of the delay line between the mixer and the tip of the electrospray needle determine the age of the reaction as the mixture enters the mass spectrometer (reaction time t_{rxn} = volume/flow rate). Thus, in the continuous flow experiment, the reaction mixture reaching the electrospray emitter is of a constant age, enabling the reaction composition at a specific time point to be investigated. In principle, this allows measurement of signal at a specific reaction age indefinitely, or at least as long as the reagent syringes contain the analyte. Adjustment of flow rate or delay loop volume can then be used to monitor different time points during the reaction.

Dramatic improvements in time resolution were achieved by Konermann et al., (52), who used a miniaturized capillary mixer system to achieve reaction times < 100 ms. As a model system, the authors used cytochrome *c*, and investigated the evolution of charge-state distributions in ESI MS resulting from folding. Native proteins tend to exhibit narrow distributions with a relatively low net charge, as a result of burial of basic sites and/or low surface area (see Section 5.1). In contrast, denatured proteins adopt significantly more extended conformations and can thus accommodate many more charges. Unfolded-state mass spectra are characterized by broad charge-state distributions centered around high charges. By acquiring data at each time point and then plotting intensity of the signal from each charge state as a function of time, one can construct a kinetic profile (Fig. 6.8). In the case of cytochrome *c*, the disappearance of the unfolded-state signal and concomitant appearance of a signal at a lower charge state corresponding to the native state, could be adequately fit to two time constants indicating, in agreement with stopped-flow spectroscopic experiments, that $\sim 90\%$ of the population folds fast with a 10% slow folding population rate limited by proline cis–trans isomerization (53).

The continuous-flow method has also been applied to study folding and reconstitution of myoglobin (39,54), and also its unfolding under acidic conditions (55). In these experiments, not only could the charge-state distributions be followed to show loss of structure, but the change in mass associated with loss of the noncovalently bound heme group also served as a probe of unfolding upon acidification. Further studies indicated that an intermediate could be detected kinetically during unfolding, but only at low pH (56), whereas at basic pH the unfolding reaction was highly cooperative.

Yet further improvements to this continuous-flow system have led to an experimental design with a dead time of $\sim$ 5 ms (57). This setup involves a primary syringe containing reactant, and an inner-capillary terminating at a mixer. This is inserted into an outer capillary containing the second reactant (e.g., refolding buffer). The two solutions meet and combine at the mixer and then pass through a short capillary reaction zone to the ESI emitter. By continuously stepping the primary syringe and mixer assembly back from the emitter, the reaction time can be increased stepwise to access a wide range of reaction times in a single continuous-flow experiment. This apparatus has been used successfully to monitor the unfolding kinetics of the dimeric protein nitric oxide synthase (58), in which the first step

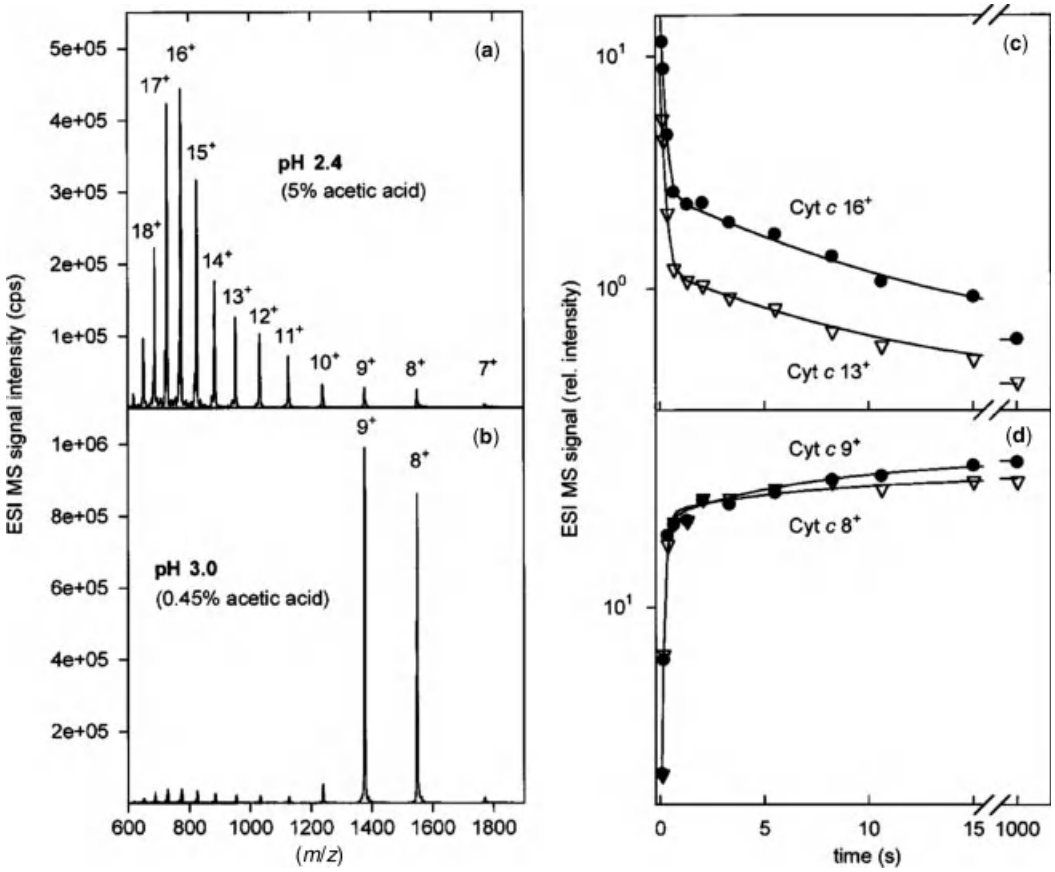


Figure 6.8. Time resolved refolding of cytochrome *c* by measuring changes in charge-state intensity. The unfolded state at pH 2.4 is characterized by maximum intensity of the 16+ charge state (*a*), whereas for the folded state the charge-state maximum is 9+ (*b*). During refolding the disappearance of the unfolded state (*c*) and appearance of native state (*d*) peaks in the mass spectrum are measured. [Reproduced with permission from (52). Copyright © 1997 American Chemical Society.]

in unfolding is acid-induced dimer dissociation, followed by rapid unfolding with concomitant dissociation of the noncovalently bound heme moiety. These processes can be monitored simultaneously by continuous-flow MS and kinetics constants determined for each (Fig. 6.9).

The HDX reactions can also be measured by continuous-flow methods (39,59). In this experimental setup, the variable delay loop is separated from the ESI needle by a second mixing chamber. Protein folding is initiated by mixing denatured protein with buffer and then passing it through a delay loop to a second mixer, where the refolding species are mixed with deuterated buffer that effects exchange at all unprotected sites. After mixing, the protein is immediately injected online into the mass spectrometer, where the extent of deuterium label incorporation is measured. The labeling time is controlled by the volume of the transfer line from this second mixer to the ESI emitter, and must be kept short to avoid exchanging amide sites that are only marginally stable. Using the short dead time experimental setup, Pan et al. (60) were able to measure folding kinetics of ubiquitin, monitoring both evolution of charge-state distributions and protection against hydrogen exchange. This allowed determination of the protection levels of each charge state to be determined at each time point, revealing the presence of folding intermediates. This online tool is powerful, since samples are analyzed immediately following the labeling pulse, which may allow hydrogen-exchange information to be retained that would be lost during sample workup.

6.2.3. Stopped-Flow Mass Spectrometry

An attractive experimental scheme for measuring protein folding or enzymatic reaction kinetics is to directly measure the folding reaction online, as can be achieved using optical spectroscopy by stopped-flow methods. After rapid mixing of reactants by turbulent flow, the reaction mixture is physically brought to a halt in the observation cell of the monitoring instrumentation. The obvious problem here is that stopped flow is, by definition, not directly compatible

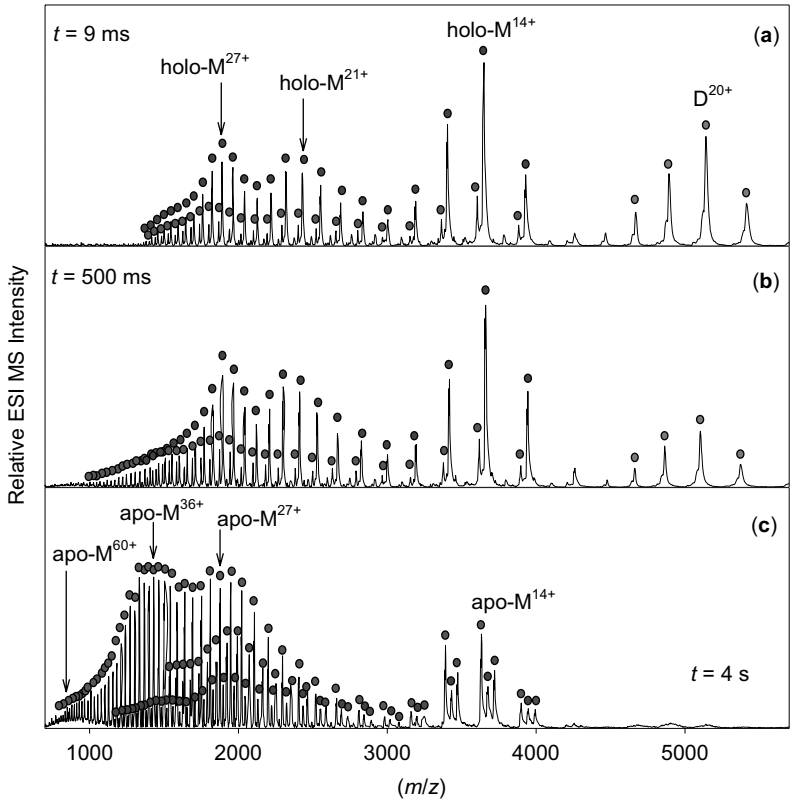


Figure 6.9. Time-resolved ESI mass spectra of nitric oxide synthase ($iNOS_{COD}$,) recorded at different times following a pH jump from 7.5 to 2.8: (*a*) $t \sim 9$ ms, (*b*) $t = 500$ ms, and (*c*) $t = 4$ s. Ions corresponding to the intact $iNOS_{COD}$ dimer, with two heme groups and two molecules of H_4B bound, are denoted with a D. Heme-bound monomers are denoted holo-M. apo-M represents monomeric proteins that are not bound to any cofactor. Also indicated are the charge states of some of the protein ions. [Reproduced with permission from (58). Copyright © 2005 American Chemical Society.]

with the continuous infusion interface required for ESI. Additionally, the very high flow rates generally used to achieve efficient mixing and shortest dead times in a stopped flow apparatus are generally incompatible with the ESI source conditions. An interesting possibility of switching flow rates has been developed for the measurement of subfemtomolar quantities of analyte using nanoflow LC–MS (61,62). With this technique, material is separated using microcapillary columns equipped with nanospray emitters. A switching valve allows the rate of infusion into the spectrometer to be switched from *high* (120 nL/min) to *low* (< 10 nL/min) flow by adjusting the split ratio of the mobile phase between column and waste. This switch effectively allows the chromatography to be paused in a data-dependent manner when a peak of interest elutes from the column, and enables the eluting compound to be analyzed for an extended period of time. This method of *peak parking* is extremely useful for species at very low concentration that may otherwise be missed or insufficiently sampled under normal flow conditions. This method has proven extremely valuable for measuring low-abundance species in proteomics, but may also find more widespread usage for kinetic studies.

Kolakowski and co-workers (63,64) circumvented these limitations by redesigning a stopped-flow setup in a manner that first allows the reaction to be rapidly initiated, and then infused into the source as a second step using a series of valves and a third syringe. In the first stage, two syringes are activated at high flow rate to mix the two reactants, to flush old solution from the reaction chamber, and then to deliver a fresh reaction mixture into the reaction tube. A valve switching system then shuts-off flow from these syringes and the reaction is delivered into the ESI source by a third syringe running at a low flow rate compatible with electrospray. Thus, relatively short dead times (100 ms) are achievable and the amount of reactant delivered corresponds to an observation time window of ~ 40 s before dilution artifacts lead to loss of signal. Careful optimization of the wash phase and solvent flow rates ensures reproducibility between stopped-flow "shots".

In this manner the authors were able to monitor, in real time using ESI MS, the acid induced unfolding of myoglobin,

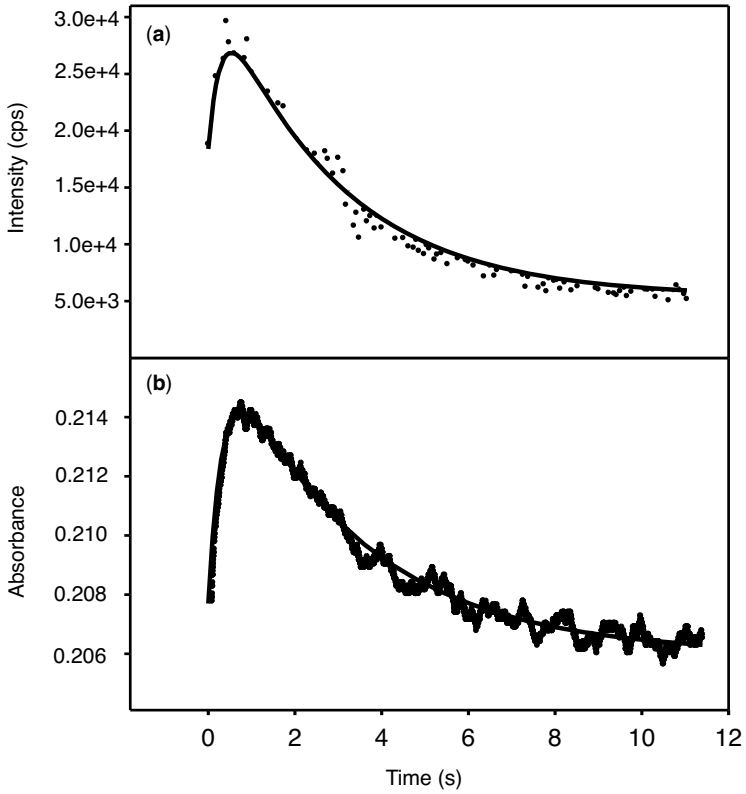


Figure 6.10. Unfolding kinetics of myoglobin at pH 3.0 monitored by stopped flow MS. The evolution of the 14+ charge state during unfolding (*a*) corresponds closely with the change in absorption of the Soret band at 441 nm (*b*). [Reproduced with permission from (64). Copyright © 2001 Elsevier Ltd.]

a process that is complete within 10 s (Fig. 6.10). Monitoring the ion intensity of the 14+ charge state of holomyoglobin resolved two kinetic phases with time constants of 0.29 and 2.8 s. These are in very good agreement with stopped-flow optical measurements, and are thought to correspond to unfolding of the holoprotein, with subsequent loss of the noncovalent heme group in the slower phase.

Although development of chip-based devices and miniature ion sources has showed some promise for the maturation of this technique (65,66), the development of ion sources that can accommodate higher flow rates and the need for improved signal-to-noise ratios (S/N) achievable only by longer time data acquisition has led to more favorable development of continuous-flow devices to achieve a similar end result.

6.2.4. Kinetics of Disulfide Formation During Folding

An alternative to HDX MS as a probe of protein folding is to monitor the change of other characteristics by MS. In the case of proteins that contain disulfide bridges, the order of their formation provides an indication of the folding pathway (67,68). As secondary and tertiary structural elements form, they orient cysteine residues correctly to enable disulfide bond formation. An elegant study by Zhang et al., (69) employed the combination of HDX MS with measurement of disulfide bond formation and evolution of charge-state distribution during the oxidative refolding of the protein human macrophage colony stimulating factor β (rhm-CSFβ). This protein forms a homodimer containing three intramolecular and three intersubunit disulfide bridges. Dilution of the reduced denatured protein into buffer containing a mixture of oxidized and reduced glutathione enabled formation of the disulfide bridges as the protein refolded. Aliquots of the refolding mixture were removed at various time points during the reaction, and any free cysteinyl residues not involved in disulfide bonds were carboxyamidomethylated by reaction with iodoacetamide. The mixture of partially refolded proteins was then separated by chromatography, digested with pepsin, and then analyzed by HPLC–ESI MS. Those fragments that contained blocked cysteine residues were identified by the mass difference resulting from alkylation. In the case of more than one cysteine occurring in the same peptic fragment, tandem MS (MS/MS) was employed to locate the modified residue. In this manner, the authors were able to determine

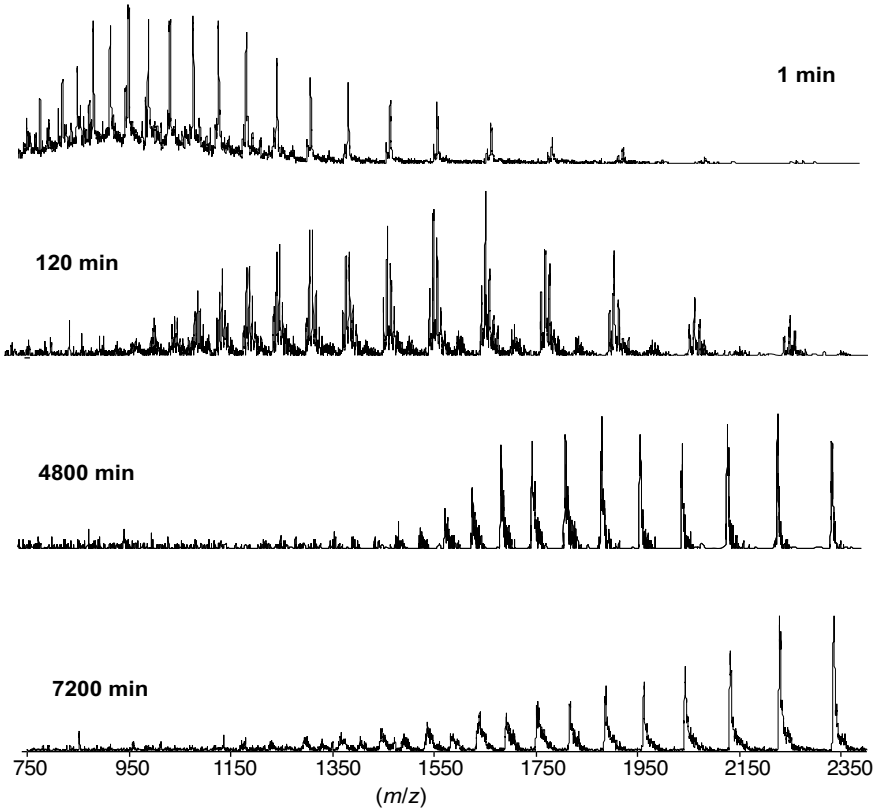


Figure 6.11. Evolution of charge-state distribution during the oxidative refolding of rhm-CSFβ. As folding proceeds and native disulfide bonds form, the protein becomes more compact, and the charge state shifts from unfolded to native-like with time. [Adapted with permission from (69). Copyright © 2002 American Chemical Society.]

not only the order of formation of structure within the monomeric subunits, but also the interactions that enable formation of the homodimer.

In addition to mapping the location of disulfide bonds described above, the partially refolded mixture was also analyzed by ESI MS, taking advantage of the exquisite sensitivity of ionic charge-state distributions to changes in the solvent-exposed surface area of the protein. At the earliest time points, only high charge states were detected, indicating the presence of predominantly unstructured chains. As oxidative refolding continued, the mass spectra became more complex due to the presence of more compact intermediate species, and eventually the formation of the dimeric species (Fig. 6.11).

A technique involving chemical labeling has also been employed in the past to study kinetics of disulfide formation during protein folding. The groups of Russell, Baldwin and their co-workers used a method termed pulsed-alkylation mass spectrometry (PA–MS) as an attractive alternative to HDX measurements (70–72). Alkylation has a significant advantage over isotope exchange reactions in that it is effectively irreversible, hence there is no possibility of back-exchange. This method relies on the specific reactivity of *N*-alkylmaleimides with free cysteine residues (i.e., those that are not involved in disulfide linkages) that are exposed to the reagent in the bulk solution. Bacterial luciferase is a 76 kDa heterodimer that contains 14 cysteinyl residues distributed throughout the α- and β- subunits, providing a set of reporters of accessibility across the structure. The authors used a pulse for 5 s with alkylating reagent followed by quenching of the reaction by addition of β-mercaptoethanol to scavenge any remaining reagent. The protein was then digested with the proteolytic enzyme AspN and the resulting fragments were identified using matrix-assisted laser desorption/ionization–time-of-flight (MALDI TOF) MS. By varying the solvent conditions, cysteine accessibility was probed either in the native state or in a partially folded species populated in 2 M urea (Fig. 6.12).

6.2.5. Irreversible Covalent Labeling As a Probe of Protein Kinetics

The above method works specifically for proteins that contain disulfide bridges. An alternative and more general approach, more akin to HDX methodology, uses the approach of labeling residues that are accessible to solvent during folding, but

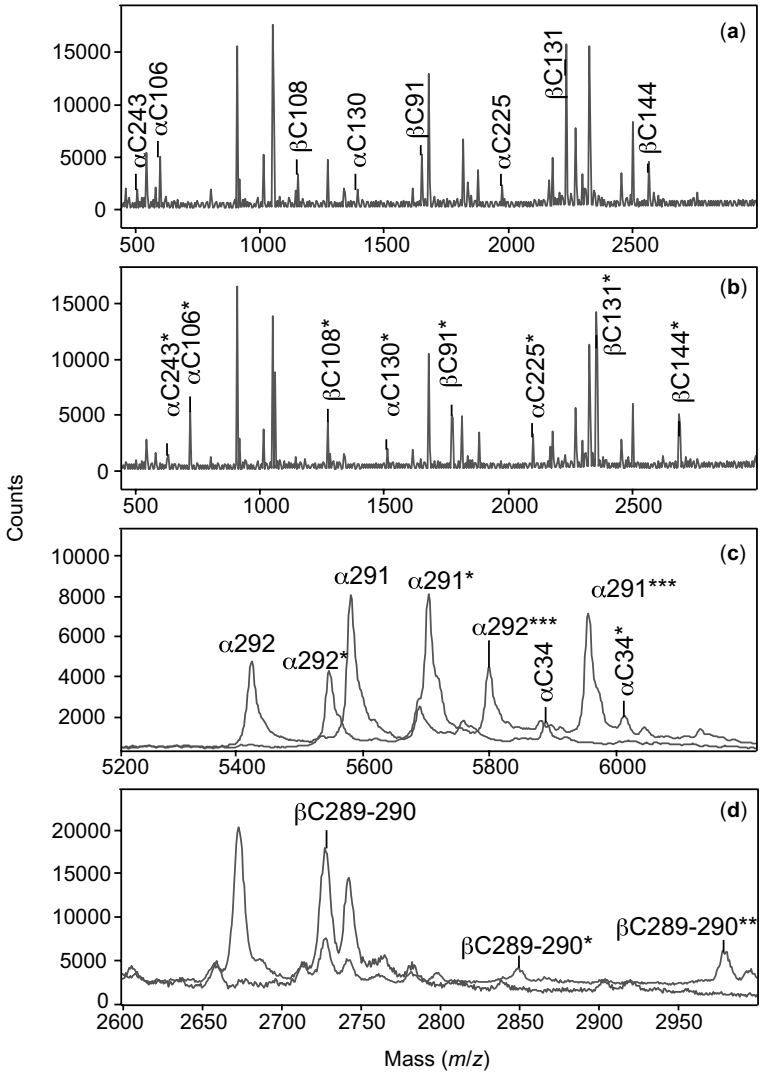


Figure 6.12. Pulsed-alkylation MS to study kinetic refolding of bacterial luciferase. As the protein refolds, disulfides are formed that prevent alkylation in folded regions. After modification the protein is digested with trypsin and the fragments analyzed by MS to determine the sites of alkylation. (*a*) Digest of unmodified luciferase, (*b*) After modification the identified peptide masses shift by 125 Da correspdonding to alkylation at a single cysteinyl residue (annotated by *). Panels *c* and *d* show superimpositions of digests before and after alkylation with the number of asterisks corresponding to the number of cysteine residues modified. [Reproduced with permission from (72). Copyright © 2001 American Chemical Society.]

with irreversible chemistry. An approach similar to that described above employs pulsed thiol labeling to probe sequestration of cysteine side chains from access to bulk solvent during folding. Site-directed mutagenesis was employed to introduce cysteine residues at 12 structural locations throughout the protein barstar (73). The time at which each position becomes buried during the folding reaction could be probed by varying when the labeling alkylation pulse was applied. This type of pulse labeling is possible using other reagents that are specific for different residue types (e.g., acetylation of lysine) (74), or have lower specificity (e.g., diethylpyrocarbonate) (75). In all cases, the key requirement is in order to obtain an accurate snapshot, labeling must be very rapid and equally rapidly quenched. Neither the reactive species nor the labeling process itself can affect the kinetics of the reaction of interest, be it protein folding or some other dynamic process.

A similar technique that has recently gained traction is covalent labeling of solvent-exposed sites by hydroxyl radicals. This again is an irreversible labeling method that obviates the need for performing further analysis under "quench" conditions, and also has the advantage of

employing a small chemical probe, that one might expect would act as a higher-resolution probe of surface accessibility than larger modifying reagents. The most popular of these methods involves oxidative labeling in a manner akin to footprinting described in previous Chapters 4 and 5. Oxidation leads to a mass difference of +16 Da that can be readily resolved for small intact proteins. Classical methods of radical formation e.g., Fenton-type reactions (76) require a metal center and produce only locally high concentrations of reactive species. A variety of innovative techniques have been developed to rapidly produce high concentrations of hydroxyl radicals, ranging from chemical sources to radical production by laser or radiolytic photolysis. These techniques have advantages over many of the other covalent labeling strategies described above, in that they are somewhat more nonspecific so there is less concern about the reactivity of individual side chains. As already discussed, there are some side chains (e.g., Cys and Met) that are nevertheless highly reactive (see Section 4.3 for a more detailed discussion of reactivity of various amino acid residues), and reactivity is somewhat modulated by the nature of neighboring residues, but the distribution of amino acids throughout the sequence and tertiary structure of the protein means that there are generally a wealth of modifiable sites. The extent of labeling can be modified simply by altering the amount of radical scavenger molecules present in the reaction mixture, making the experimental setup as part of a continuous-flow device compatible with online measurements by MS. Examples of this technique in action to study the kinetics and folding and unfolding of proteins are described below.

The Gross (77,78) laboratory termed their design "fast photochemical oxidation of proteins (FPOP)". This approach involves flash photolysis of hydrogen peroxide with a KrF excimer laser at 248 nm to form $HO^{\bullet}$ radicals in high concentration. Its appeal for kinetic studies stems from the fact that the lifetime of hydroxyl radicals is restricted to $\sim 1\,\mu$s by the presence of a radical scavenger (e.g., glutamine). Any protein side chains that are exposed to solvent, however, have a high chance of becoming oxidized, forming a enduring modification that acts as an indicator of its exposure at that time point during protein folding. If the excitation is pulsed at various times after the initiation of protein refolding, then the labeling reaction acts as a probe of protein conformation that can be used to build up a picture of the topology of the folding reaction, subject to the limitations of side chain reactivity mentioned above. Alternatives to peroxide have also been investigated that may give different modification specificity, such as the sulfate radical anion ($SO_4^{\bullet-}$ generated by photolysis of persulfate), which appears to be more reactive and thus may yield more rapid or more extensive modification products (79).

A recent extension to this technique was designed to bring even earlier protein-folding events into the detection window of the FPOP method. Using a second laser emitting in the infrared (IR) region of the spectrum (1900 nm), a very rapid temperature jump can be achieved by heating the aqueous solvent. Using barstar, a small protein that is denatured at low temperature, a very short IR laser pulse of only a few nanoseconds was sufficient to induce a 20°C temperature jump in the reaction volume and to induce protein refolding (Fig. 6.13). Subsequent photolysis of endogenous peroxide in the refolding mixture at various time points after initiation of folding enabled the probing of submillisecond protein folding events (80). Although in this study only the extent of oxidation of the whole protein was monitored, in principle this could be extended to localize sites of modification using standard top-down or bottom-up approaches.

Similar experiments have been performed in the Konermann laboratory, who have investigated the folding of cytochrome *c* by pulsed oxidative labeling, again employing laser flash photolysis of H_2O_2 as the source of hydroxyl radicals (81). Short-lived structural intermediate species were detected during the kinetic unfolding of myoglobin by this methodology, following the pulsed oxidation step with bottom-up analysis of tryptic digests. By measuring the sites of oxidation as a function of time, the regions that unfold at various time points and order of unfolding could be determined (82). More recently, this approach has been extended to more complex systems including measuring the folding and assembly pathway of a 22 kDa dimeric protein S100A11 (83), and even to address the folding of the integral membrane protein rhodopsin in bicelles with and without the presence of the cofactor retinal (84). Hydroxyl radicals have advantages over some other labeling reagents in that they may penetrate into the detergent bicelles, thus potentially probing regions that would otherwise be inaccessible. Clearly, the diversity of systems that have already been studied in a relatively short period of time suggests that the versatility of this technique has still to be fully exploited. For further details of the technique and its applications to structural and dynamic studies of proteins, the reader is directed to a number of excellent recent reviews (85–88).

6.3. KINETICS OF PROTEIN ASSEMBLY

With the advent of instrumentation capable not only of retaining the structure of macromolecular complexes and transferring them into the gas phase (ESI and nano-ESI), but also of detecting large complexes that manifest at high *m/z* (particularly hybrid ESI Q–TOF), there has been substantial interest in obtaining structural information about large macromolecules and molecular machines, as will be discussed further in Section 9.4. There has been significantly

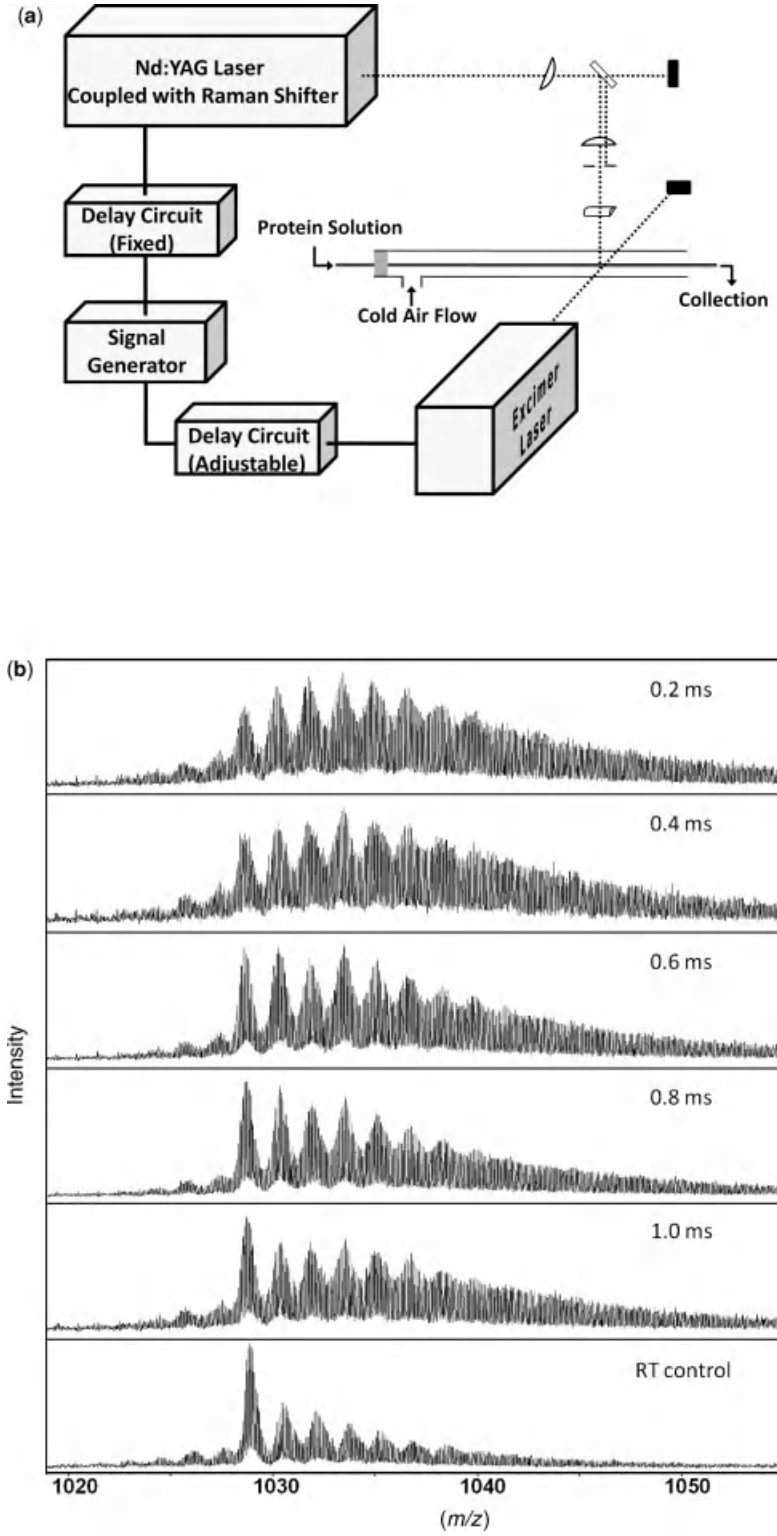


Figure 6.13. (*a*) Schematic of the FPOP experimental flow system intersected by two laser beams at a window in the tube. The time between the two laser pulses is adjustable with the "delay circuit". (*b*) Representative mass spectra of the barstar post-FPOP as a function of the time between the heating pulse and the FPOP probe. [Reproduced with permission from (80). Copyright © 2010 American Chemical Society.]

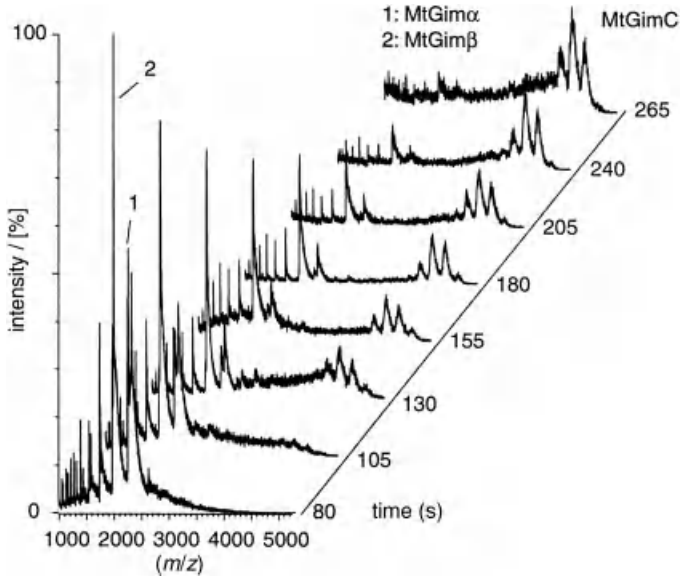


Figure 6.14. Assembly kinetics in MtGimC from subunits MtGimα and MtGimβ. As time progresses peaks arise from the intact complex and the monomer becomes depleted. [Reproduced with permission from (89). Copyright © 2000 National Academy of Sciences, USA.]

less focus, however, on the kinetic processes of assembly and disassembly of these species, whether they be homo-oligomeric proteins (e.g., the GroEL chaperone) or hetero-multimers (e.g., ribosome). For instance, the chaperone GroEL is a 60 kDa protein, but the active form requires oligomerization to form two seven-membered rings stacked on top of each other, forming an 800 kDa homotetradecamer. The kinetics and mechanism by which proteins assemble to form multimeric complexes that represent the active form is an important extension of the protein folding field.

In the simplest case, the assembly process can be monitored in real time by adding the components together in solution and continuously infusing the sample into the mass spectrometer. As proteins form associations in solution, their change in molecular mass can be directly read out from the evolution of the mass spectrum. In practice, it may be difficult to isolate the individual components initially, so valuable information can be obtained from perturbation of preformed complexes by instead measuring complex dissociation. These types of experiments were pioneered by the Robinson group and have provided valuable insight into the assembly processes of large multiprotein complexes. One of the first examples of kinetic measurement of the assembly process was a study of the chaperone complex MtGimC (89). The active form of this protein is a heterohexamer ($\alpha_2\beta_4$) formed from two subunits, MtGimα and MtGimβ. The intact hexamer is 87 kDa and the charge states adopted by the native state complex give rise to peaks in the range 4000–5000 *m/z*. Using very gentle source conditions, including elevated pressures within both the source and analyzer regions of an ESI–TOF mass spectrometer, the authors were able to observe the intact complex and various partially dissociated species. Additionally, manual mixing of the α- and β- subunits in a 1:2 ratio at 37°C followed by infusion into the mass spectrometer allowed the assembly process to be monitored in real time. Interestingly, the assembly process appeared to be a highly cooperative event involving no observable intermediates (Fig. 6.14). Over the time course measured, only peaks corresponding to the intact hexamer were detected, with consequent depletion of the monomer populations. In contrast, controlled dissociation revealed a variety of dimers, trimers, and tetramers, all consistent with a $(\beta\alpha\beta)_2$ architecture in which the trimers appear to be linked primarily through the α-subunit.

An interesting example of protein disassembly under mild conditions suggests that observed intermediates might represent evolutionary intermediates. For instance, the hexameric adenosine triphosphate (ATP) sulfurylase from *Penicillium chrysogenum* is predicted to have evolved from a dimeric ancestor, and these are, indeed, the dominant species observed upon mild dissociation in the mass spectrometer, rather than alternative possibilities (e.g., trimer or tetramer) (90). Assembly of the 20S proteasome has also been investigated by real-time MS. This molecular machine contains α- and β- subunits assembled to form a 28-mer complex. Mixing the monomer units and observing the evolution of mass spectral signals enabled observation of the process occurring via formation of $\alpha\beta$ heterodimers, the $\alpha_7\beta_7$ 14-mer, and finally the intact $\alpha_{14}\beta_{14}$ proteasome over the time course of 30 min (91).

Another intriguing study of protein assembly by MS involves the kinetics of subunit exchange in small heat-shock proteins (sHSPs). This diverse family of proteins has a common feature that, under heat-shock conditions, structural rearrangements occur such that the oligomeric species can bind to unfolded or partially folded proteins and protect them against aggregation. Early studies demonstrated that the oligomeric state is highly dynamic, with polydisperse oligomers forming that range in size from 9 to 24 monomer subunits. The heterogeneous nature leads to difficulties in crystallographic structure determination, or even characterization by electrophoresis. Once again the power of MS is evident: using nano-ESI and measurement by TOF MS, the polydispersity and dynamic nature of these species can be probed. An artificial mixture of homologous sHSPs obtained from different species (pea and wheat) were shown to interact and form dynamic dodecamers (Fig. 6.15) By observing the kinetics of subunit exchange (by virtue of the mass difference between the homologous proteins), the authors were able to determine that, although exchange is relatively facile, the predominant exchanging species is a dimer (92). A similar observation was made for subunit exchange between pea sHSP18.1 and sHSP17.6, which exist

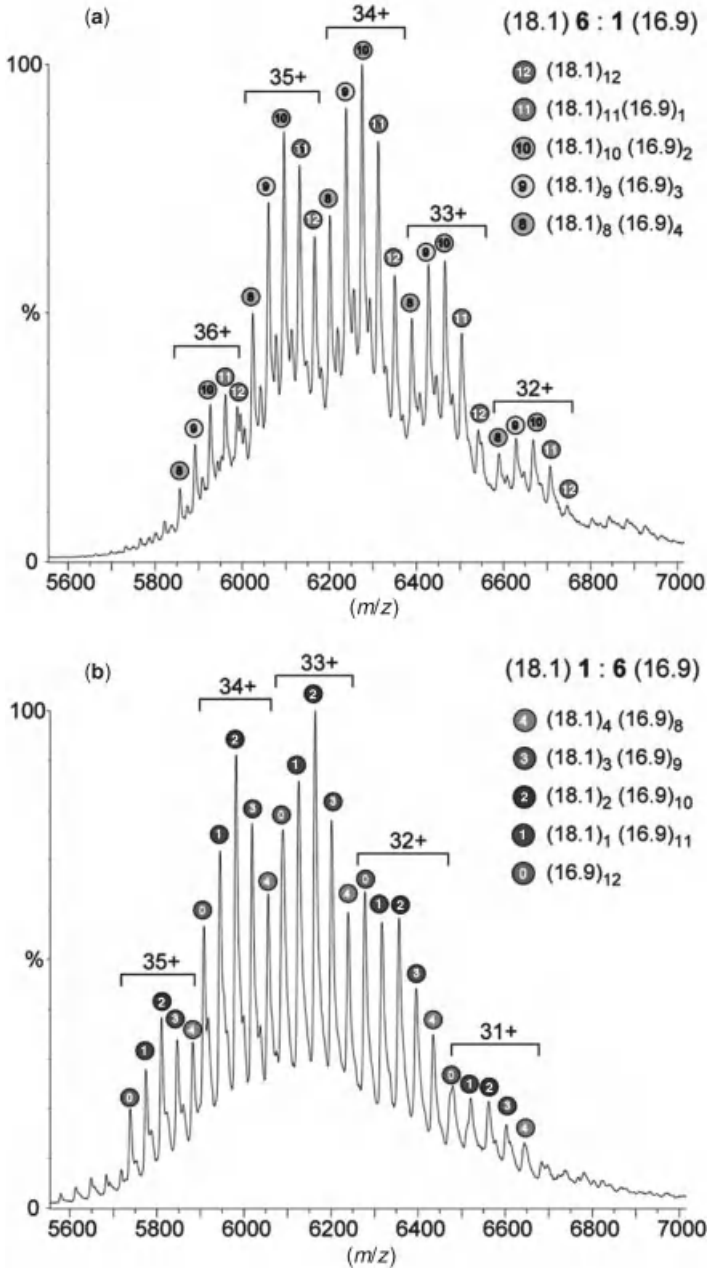


Figure 6.15. Subunit exchange in the sHSPs. Mixing PsHSP18.1 from pea with TaHSP16.9 from wheat in different ratios allows subunit exchange between the two to occur, producing a variety of different stoichiometries over time that can be characterized by MS. [Reproduced with permission from (92). Copyright © 2002 American Society for Biochemistry and Molecular Biology.]

predominantly as a dodecamer at physiological temperatures, but are able to exchange dimer subunits on the time scale of minutes (93). Interestingly, under heat-shock conditions, rearrangement occurs to form larger oligomers that are highly dynamic in nature, and the dynamic change in mass of the various complexes could be observed directly by ESI MS (Fig. 6.16) (94,95). This approach was extended to measure kinetics of association of the chaperone oligomer with a client protein luciferase, and since the resulting spectra are complicated by heterogeneity, identification of the many species present could be made by dissociating individual complexes by MS/MS. Such structural plasticity and reorganizational ability is presumably vital for the chaperone action of these proteins. A similar approach of monitoring exchange subunits was taken recently by Chevreux et al. (96), who mixed ^{14}N- and ^{15}N- labeled subunits

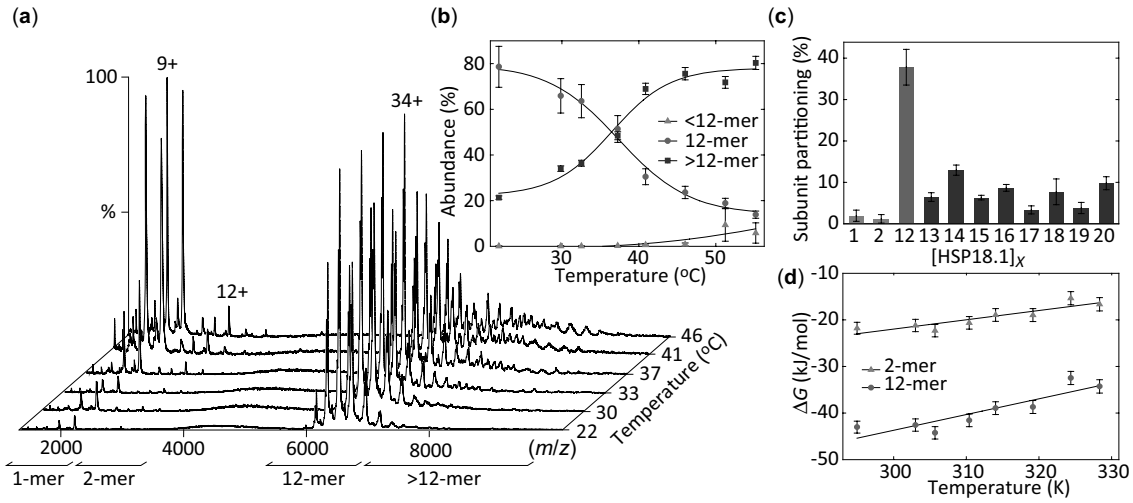


Figure 6.16. Temperature-induced changes in HSP18.1 oligomerization. (*a*) nano-ESI mass spectra of HSP18.1 obtained at temperatures from 22 to 46°C. (*b*) Plot of the relative amount of HSP18.1 subunits existing in different oligomeric states. (*c*) Abundance of the different species at 46°C shows a clear preference for oligomers with an even number of subunits. (*d*) Differences in free energy between a subunit free in solution or incorporated into either a dimer or dodecamer follow linear trends, suggesting that neither undergo significant structural change upon thermal activation. [Reproduced with permission from (94). Copyright © 2010 National Academy of Sciences, USA.]

of the protein glucosamine-6-phosphate synthase. By using the changes in the mass of the complex, they were able to monitor the kinetics of exchange of subunits in the dimer.

Real-time observation of protein complex assembly and disassembly has been largely facilitated by the innovation of automated microfluidics-based platforms for nano-ESI, enabling highly reproducible results to be obtained without the uncertainty of sample handling and human errors. Systems, such as those developed by Henion and coworkers (97) have revolutionized the field, and are now commercially available. The ability to mix controlled small volumes of reactant in a temperature-controlled environment and to deliver them to nano-ESI emitters, coupled with the availability of reproducible emitters, makes it possible to perform experiments under varying experimental conditions, that are free of instrumental artifacts that could affect the outcome.

A relative newcomer to the field of kinetic protein assembly is ion mobility MS (IM–MS). This method has the advantage of being able to separate species present in complex mixtures based on their physical shape in the gas phase and, hence, drift time. It has potential use for simplifying the analysis of a complex mass spectrum containing proteins of different mass, but is absolutely invaluable for the differentiation of isobaric species that are conformationally different. Two recent studies used IM–MS for the study of amyloid formation. In one case, the assembly of amyloid protein Aβ42 into oligomers was monitored using ion mobility to separate oligomeric species based on their shapes (98). From the combination of mass measurement to determine the oligomeric state and drift time to infer the organization of the oligomers (linear versus cyclic), a mechanistic model of assembly could be suggested. In another elegant study, subunit exchange was measured between ^{14}N- and ^{15}N-labeled β_2-microglobulin fibrils (99). Although exchange occurred quite rapidly in small oligomers, a difference in kinetics was observed between trimers with different ion mobilities. This suggested that two trimer structures exist, a more compact form with slow exchange kinetics and a more dynamic extended species.

In this section, we highlighted just a few of the many protein assembly studies that have been performed to date. Investigation of protein assembly, stoichiometry, and dynamics is a rapidly evolving field in which MS is clearly poised to play a very important role.

6.4. KINETICS OF ENZYME CATALYSIS

In Section 6.3, we saw some of the mass spectrometric methods used to investigate the folding and unfolding reactions of proteins. Armed with this information, we can gain insight into how a protein achieves its native folded and functional structure. Equally important though is an understanding of how the folded protein functions. In Chapter 5, the importance of dynamics was discussed, and the many ways we can look at the dynamic events occurring in the native structure. One particular class of proteins is also involved in the catalysis of complex chemical reactions; namely, enzymes are capable of transforming reactants into

products in a highly efficient and specific manner. Biophysicists are extremely interested in the kinetics of these reactions and how nature has optimized the structure and function of proteins to best achieve their genetically destined goals. This section briefly discusses some of the mass spectrometric approaches to study the kinetics of enzyme catalysis.

Enzymes are extremely efficient biological machines, converting substrates to products on time scales similar, or often faster, than the process of protein folding. Their action is catalytic, so they are capable of performing multiple turnovers provided that there are sufficient reactants present. For this reason, it is often easier to obtain at least limited kinetic information about the reaction process, since it is continuous. In the case of protein folding, once the denatured protein is exposed to conditions that are permissive for folding, the entire population proceeds to fold toward the native state, albeit potentially at different rates depending on the pathway(s) available. For enzyme catalysis, one can employ a limiting concentration of catalyst in the presence of excess substrate and observe the rate of reaction.

The simplest method to measure enzyme kinetics is at the so-called "steady state". In an excess of substrate, a rapid equilibrium is set up in which the concentration of intermediate species builds up until they are effectively at a steady state, being formed and depleted at the same rate. Substrate is gradually depleted over time, but is present in such excess that, for a sufficiently short-time window of measurement, this too can be considered constant. Under such conditions, the reaction rate is more or less constant so that the rate of formation of products and the dissociation constant for enzyme–substrate complex can be measured.

The theory of enzyme catalysis was developed and proposed by Michaelis and Menten (100), based on the formation of a noncovalent complex of enzyme (E) and substrate (S), at equilibrium, from which the reaction is catalyzed to yield product (P). The general reaction scheme may be represented as:

$$\mathrm{E} + \mathrm{S} \overset{K_s}{\rightleftharpoons} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P} \qquad (6\text{-}4\text{-}1)$$

assuming substrate is present in vast excess over the enzyme concentration. The equation $K_S = [ES]/[E][S]$ represents the dissociation constant for the enzyme–substrate complex (also known as K_m, the Michaelis constant). The chemical reaction that converts reactants to products then has a reaction velocity $v = k_{cat}[ES]$, which leads to the classic Michaelis–Menten equation:

$$v = \frac{[E]_0[S]k_{cat}}{K_m + [S]} \qquad (6\text{-}4\text{-}2)$$

This equation is valid with the assumption that substrate is present in vast excess over the enzyme. Generally, kinetic experiments are performed under conditions where the substrate concentration is saturating, leading to a limiting maximum reaction velocity, $V_{max} = k_{cat}[E]_0$.

Traditionally, measurements of steady-state kinetics were achieved spectroscopically using a chromogenic substrate that undergoes a color change, thus enabling different parts of the reaction to be probed using different artificial substrates (101). However, this method is generally not applicable to natural substrates and is thus necessarily limited in scope. Another problem with modified substrates is that the very presence of the chromophore may alter the kinetics of catalysis (102). For this reason, there has been a great deal of interest in using mass spectrometric methods for the analysis of enzyme kinetics, since it effectively serves as a "universal" detector (103). One of the great powers of MS is that, with a sufficient time resolution, not only can kinetics be monitored as the change in intensity of spectral peaks, but also the intermediate species present can be determined, provided the reaction gives rise to a resolvable change in mass. If the catalytic reaction occurs within a time scale amenable to measurement online then one can simultaneously monitor the kinetics of formation of products and infer mechanistic details from the detection of intermediate species.

One of the earliest direct measurements of enzyme catalysis was described by Smith and Caprioli (104,105), who used fast atom bombardment (FAB) ionization to measure the hydrolysis of peptides by trypsin. This method, however, requires a high concentration of glycerol as liquid matrix that can rapidly cause inactivation of enzymes. Electrospray ionization provides a much more friendly solvent environment, as shown by Lee et al. (49), who employed continuous-flow MS to measure the kinetics of a variety of enzymatic reactions, including the hydrolysis of *O*-nitrophenyl β-D-galactopyranoside by lactase and the proteolytic degradation of a short peptide, dynorphin 1–8, by α-chymotrypsin. These reactions occurred on a sufficiently slow time scale, such that manual mixing followed by infusion into the mass spectrometer was sufficient to enable reaction kinetics to be measured directly as depletion of reactant and formation of product. Quantitation of kinetic parameters can be achieved by use of internal standards, and by separating various species by HPLC. This avoids the potential problems of suppression of ionization that would give rise to nonlinear concentration responses. Henion and co-workers (106) demonstrated the applicability of HPLC–MS to the measurement of kinetics of hydrolysis by β-galactosidase of its cognate substrate lactose, using a disaccharide as an internal standard for quantitation.

Steady-state kinetic measurement is a useful but limited technique. While it can be used to directly measure the overall rate of reaction (k_{cat}) and the Michaelis constant

(K_m) for the reaction, it reveals very little mechanistic detail about the reaction. Only by determining the rates of formation and depletion of intermediate states can a full description of the enzymatic reaction be made. Using MS, it has proven possible to detect both covalently and noncovalently bound intermediate enzyme–substrate complexes that, when identified, can provide quite detailed mechanistic information. In addition, with rapid mixing and detection, determining the rates of the processes that interconvert these intermediate species allows a more complete kinetic description of the processes.

As with studies of protein folding, pushing the time scales shorter and shorter can provide more useful details about enzyme catalysis. For instance, the observation of transient intermediate species can provide insight into the mechanism of an enzymatic reaction and the species that become populated during the process. Additionally, measurement of the rates of formation of these species enables a more complete kinetic picture to be drawn. Various approaches have been taken to address these questions using MS, by employing many of the methodologies similar to those described above for protein folding studies. Chemical quench techniques followed by off-line detection using radiolabeling techniques or NMR, for instance, can be used to detect intermediate species, but only if they are stable to the quench conditions employed (107–109). However, even in the case where the transient intermediates are labile they may be detectable using MS. *Pulsed-flow* mixing with millisecond resolution enabled detection of a tetrahedral intermediate in the conversion of shikimate-3-phosphate to phosphoenol pyruvate by the enzyme 5-enoylpyruvyl-shikimate-3-phosphate synthase by ESI MS (110). This setup was effectively the same as a continuous flow mixing device that was able to achieve a time resolution on the order of 50 ms. Although no kinetic measurements were made, a number of peaks corresponding to intermediates in the reaction pathway of this enzyme were observed without the need to quench the reaction before detection.

An elegant example of the application of MS to the study of much more rapid enzyme kinetics was given by Li et al. (111), who used a continuous-flow mixing device with variable flow rates to monitor the reaction of 3-deoxy-D-manno-2-octulosonate-8-phosphate (KDO8P) synthase. This enzyme produces KDO8P by condensation of arabinose-5-phosphate (A5P) with phosphoenol pyruvate (PEP). Using ESI MS and flow rates that allowed a time resolution from 7 to 160 ms, the authors were able to detect a variety of enzyme complexes not only with the substrate and the product, but also with an intermediate species that had previously only been postulated. Since each of these species differs in mass, they could all readily be observed and resolved from each other. With improved time resolution, it is possible to extract kinetic constants and thus define the catalytic reaction of this enzyme in more detail. The beauty of this technique is the ability not only to monitor concentrations of reactant and product moieties, but also to detect any intermediate species present as long as the reaction involves a change in the molecular weight.

Continuous-flow MS may also be used to obtain kinetic constants of an enzymatic reaction. If the time resolution of the mixer and the interface to the mass spectrometer is sufficient, then it becomes possible to measure directly the rates of formation of enzyme intermediate species. Pre-steady state kinetic measurements were performed on a single point mutant of the protein xylanase, which hydrolyses β-1,4-glycosidic bonds in xylan to form disaccharides, at a rate sufficiently slow to resolve transient glycosyl enzyme intermediate species by continuous-flow MS. With the use of an artificial substrate 2,5-dinitrophenyl β-xylobioside (2,5-DNPX_2), the authors were able to follow the hydrolysis reaction photometrically from the release of 2,5-dinitrophenol (2,5-DNP), and also mass spectrometrically by observation of the intermediate E-X_2 (112). Another example employed the dual capillary system described in Section 6.2.2 (57) to investigate in real time the proteolysis of bradykinin by chymotrypsin (113).

Other studies have demonstrated the power of MS for the identification of intermediate species and for mechanistic clarification. One such study of TEM-2 β-lactamase, a well studied enzyme, and its inhibition by clavulanic acid, called into question the mechanism of inhibition proposed by previous research (114). Previous studies had suggested that a single serine (Ser^{70}) was responsible for binding the carboxylate group of the inhibitor, with subsequent opening of the β-lactam ring. By observating time course inhibition, proteolytic digestion of the enzyme–inhibitor complex, and subsequent mass analysis of the digest products, the authors were able to show that not only was Ser^{70} an important binding residue, but also a second site, Ser^{130}, was involved in cross-linking to the inhibitor during the reaction.

A number of enzymatic processes have been described recently whose mechanistic details were elucidated by MS (115–118). The use of single-ion monitoring (SIM) of substrate and/or product provides a high-sensitivity method of quantitation. In some cases, the kinetic parameters obtained led to mechanistic hypotheses. A particularly elegant example is a study on the assembly during biosynthesis of the natural product antibiotic novobiocin. This process requires four different enzymes that act in sequence to create the tripartite scaffold of the final product. Simultaneous quantitative measurement of the products of each enzymatic reaction enabled the authors to build up a picture of this complex assembly process (119), with the goal of a robust assay that could be used to assess whether alternate substrates could be employed to produce novel antibiotics in this structural class.

MALDI TOF MS has become quite a routine tool for the measurement of enzyme catalysis, and has also been used for

the investigation of pre-steady state kinetics. For example, quantitation of a phosphoenzyme intermediate as a function of time during the dephosphorylation reaction of phosphotyrosine by protein tyrosine phosphatase provides both kinetic information and mechanistic insight (120). A recent exciting development in this field combines microfluidics with laser ionization MS. Miniaturization of solvent flow paths and an array of addressable electrodes enable single solvent droplets containing reactants to be mixed, allowed to react, and then quenched by a third droplet and mixed with MALDI matrix (121). Alternatively, using a microfluidic reactor fashioned from silicon, this in itself can be used directly for laser desorption of small-molecule substrates and products [desorption ionization on silicon, DIOS MS (122)]. Reaction kinetics of the conversion by the enzyme arginase of arginine to ornithine were monitored by mixing reactants and then, as the reaction mixture passed along the length of a microfluidic channel, it deposited a residue whose contents represented the substrate–product concentrations at that time point. Reaction kinetics could be determined (121,123) by measuring the DIOS MS signal for the substrate along the length of the channel.

In the pharmaceutical industry and elsewhere, there is increased emphasis on high-throughput applications and this is especially true in the field of enzyme kinetics. Of particular interest is the ability to screen a wide range of inhibitors against an enzyme target. The use of online desalting in a microfluidic device and SIM measurement of substrate and/or product gives sufficiently high sensitivity for a rapid assay that can be used to very quickly screen compounds (124). The development of these MS-based enzyme assays and adaptation to use in a high-throughput environment is described in a recent review (125).

This chapter showed how kinetic measurements can be employed to elucidate mechanistic details of protein folding and function. Coupled with studies of dynamics at equilibrium in Chapter 5, this shows how MS is becoming an extremely powerful tool in the biophysical arena.

REFERENCES

1. Roughton, F. J. W. (1934) The kinetics of haemoglobin IV–general methods and theoretical basis for the reactions with carbon monoxide, *Proc. R. Soc. London, Ser. B: Bio. Sci. 115*, 451–464.
2. Chance, B. (1940) The accelerated-flow method for rapid reactions, *J. Franklin Inst. 229*, 455–766.
3. Shastry, M. C. R., Luck, S. D., and Roder, H. (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale, *Biophys. J. 74*, 2714–2721.
4. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. (1983) *Molecular Biology of the Cell*, Grandland Publishing, Inc., New York.
5. Dyson, R. (1974) *Cell Biology*, Allyn and Bacon, Inc., Boston.
6. Darnell, J., Lodish, H., and Baltimore, D. (1990) *Molecular Cell Biology*, W.H. Freeman and Company, New York.
7. Jacobs, M. D. and Fox, R. O. (1994) Protein folding intermediates characterized by pulsed hydrogen exchange, *Tech. Protein Chem. 5*, 447–454.
8. Roder, H., Elöve, G. A., and Englander, S. W. (1988) Structural characterisation of folding intermediates in cytochrome *c* by hydrogen exchange labelling and proton NMR, *Nature (London) 335*, 700–704.
9. Varley, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H., and Clore, G. M. (1993) Kinetics of Folding of the All b-Sheet Protein Interleukin-1-β, *Science 260*, 1110–1113.
10. Radford, S. E., Dobson, C. M., and Evans, P. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature (London) 358*, 302–307.
11. Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., and Dobson, C. M. (1993) Detection of transient protein folding populations by mass spectrometry, *Science 262*, 896–900.
12. Katta, V. and Chait, B. T. (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 5*, 214–217.
13. Hooke, S. D., Eyles, S. J., Miranker, A., Radford, S. E., Robinson, C. V., and Dobson, C. M. (1995) Cooperative elements in protein folding monitored by electrospray ionisation mass spectrometry, *J. Am. Chem. Soc. 117*, 7548–7549.
14. Matthews, C. R. (1993) Pathways of protein folding, *Ann. Rev. Biochem. 62*, 653–683.
15. Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure, *Proteins 6*, 87–103.
16. Baldwin, R. L. (1994) Finding intermediates in protein folding, *Bioessays 16*, 207–210.
17. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native state hydrogen exchange, *Science 269*, 192–196.
18. Zitzewitz, J. A. and Matthews, C. R. (1993) Protein engineering strategies in examining protein folding intermediates, *Curr. Opin. Struct. Biol. 3*, 594–600.
19. Dill, K. A. and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10–19.
20. Wolynes, P. G., Luthey-Schulten, Z., and Onuchic, J. N. (1996) Fast-folding experiments and the topography of protein folding energy landscapes, *Chem. Biol. 3*, 425–432.
21. Jackson, S. E. and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry 30*, 10428–10433.
22. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Folding Design 3*, R81–R91.

23. Kragelund, B. B., Hojrup, P., Jensen, M. S., Schjerling, C. K., Juul, E., Knudsen, J., and Poulsen, F. M. (1996) Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family, *J. Mol. Biol. 256*, 187–200.
24. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C., and Radford, S. E. (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9, *J. Mol. Biol. 286*, 1597–1608.
25. Gunasekaran, K., Eyles, S. J., Hagler, A. T., and Gierasch, L. M. (2001) Keeping it in the family: folding studies of related proteins, *Curr. Opin. Struct. Biol. 11*, 83–93.
26. Kiefhaber, T. (1995) Kinetic traps in lysozyme folding, *Proc. Natl. Acad. Sci. USA 92*, 9029–9033.
27. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) The barriers in protein folding, *Nat. Struct. Biol. 1*, 149–156.
28. Heidary, D. K., Gross, L. A., Roy, M., and Jennings, P. A. (1997) Evidence for an obligatory intermediate in the folding of interleukin 1β, *Nat. Struct. Biol. 4*, 726–731.
29. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate, *Protein Sci. 8*, 45–49.
30. Nishimura, C., Prytulla, S., Dyson, H. J., and Wright, P. E. (2000) Conservation of folding pathways in evolutionarily distant globin sequences, *Nat. Struct. Biol. 7*, 679–686.
31. Nishimura, C. Wright, P. E., and Dyson, H. J. (2003) Role of the B helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions, *J. Mol. Biol. 334*, 293–307.
32. Jennings, P. A. and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, *Science 262*, 892–896.
33. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate, *Science 249*, 1544–1548.
34. Cavagnero, S., Theriault, Y., Narula, S. S., Dyson, H. J., and Wright, P. E. (2000) Amide proton hydrogen exchange rates for sperm whale myoglobin obtained from ^{15}N-^{1}H NMR spectra, *Protein Sci. 9*, 186–193.
35. Nishimura, C., Dyson, H. J., and Wright, P. E. (2002) The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate, *J. Mol. Biol. 322*, 483–489.
36. Bieri, O. and Kiefhaber, T. (2001) Origin of apparent fast and non-exponential kinetics of lysozyme folding measured in pulsed hydrogen exchange experiments, *J. Mol. Biol. 310*, 919–935.
37. Deng, Y., Zhang, Z., and Smith, D. L. (1999) Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics, *J. Am. Soc. Mass Spectrom. 10*, 675–684.
38. Udgaonkar, J. B. and Baldwin, R. L. (1990) Early folding intermediate of ribonuclease A, *Proc. Natl. Acad. Sci. USA 87*, 8197–8201.
39. Simmons, D. A., and Konermann, L. (2002) Characterization of transient protein folding intermediates during myoglobin reconstitution by time-resolved electrospray mass spectrometry with on-line isotopic pulse labeling, *Biochemistry 41*, 1906–1914.
40. Yang, H. and Smith, D. L. (1997) Kinetics of cytochrome *c* folding examined by hydrogen exchange and mass spectrometry., *Biochemistry 36*, 14992–14999.
41. Pan, H. and Smith, D. L. (2003) Quaternary structure of aldolase leads to differences in its unfolding and unfolding intermediates, *Biochemistry 42*, 5713–5721.
42. Wolynes, P. G. (1997) Folding funnels and energy landscapes of larger proteins within the capillary approximation, *Proc. Natl. Acad. Sci. USA 94*, 6170–6175.
43. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis, *Proteins 21*, 167–195.
44. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2008) Electron capture dissociation of electrosprayed protein ions for spatially resolved hydrogen exchange measurements, *J. Am. Chem. Soc. 130*, 11574–11575.
45. Rand, K. D., Pringle, S. D., Murphy, J. P., Fadgen, K. E., Brown, J., and Engen, J. R. (2009) Gas-phase hydrogen/deuterium exchange in a traveling wave ion guide for the examination of protein conformations, *Anal. Chem. 81*, 10019–10028.
46. Pan, J. X., Han, J., Borchers, C. H., and Konermann, L. (2009) Hydrogen/deuterium exchange mass spectrometry with top-down electron capture dissociation for characterizing structural transitions of a 17 kDa protein, *J. Am. Chem. Soc. 131*, 12801–12808.
47. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2010) Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry, *Anal. Chem. 82*, 8591–8597.
48. Kaltashov, I. A., Bobst, C. E., and Abzalimov, R. R. (2009) H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach?, *Anal. Chem. 81*, 7892–7899.
49. Lee, E. D., Muck, W., Henion, J. D., and Covey, T. R. (1989) Real time reaction monitoring by continuous-introduction ion spray tandem mass spectrometry, *J. Am. Chem. Soc. 111*, 4600–4604.
50. Sam, J. W., Tang, X.-J., and Peisach, J. (1994) Electrospray mass spectrometry of iron bleomycin: demonstration that activated bleomycin is a ferric peroxide complex, *J. Am. Chem. Soc. 116*, 5250–5256.
51. Sam, J. W., Tang, X.-J., Magliozzo, R. S., and Peisach, J. (1995) Electrospray mass spectrometry of iron bleomycin II: investigation of the reaction of Fe(III)-bleomycin with iodosylbenzene, *J. Am. Chem. Soc. 117*, 1012–1018.
52. Konermann, L., Collings, B. A., and Douglas, D. J. (1997) Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry, *Biochemistry 36*, 5554–5559.
53. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein

denaturation reactions is due to cis-trans isomerism of proline residues., *Biochemistry 14*, 4953–4963.

54. Lee, V. W. S., Chen, Y.-L., and Konermann, L. (1999) Reconstitution of acid-denatured holomyoglobin studied by time resolved electrospray ionization mass spectrometry, *Anal. Chem. 71*, 4154–4159.
55. Konermann, L., Rosell, F. I., Mauk, A. G., and Douglas, D. J. (1997) Acid-induced denaturation of myoglobin studied by time-resolved electrospray ionization mass spectrometry, *Biochemistry 36*, 6448–6454.
56. Sogbein, O. O., Simmons, D. A., and Konermann, L. (2000) Effects of pH on the kinetic reaction mechanism of myoglobin unfolding studied by time-resolved electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 11*, 312–319.
57. Wilson, D. J. and Konermann, L. (2003) A capillary mixer with adjustable reaction chamber volume for millisecond time-resolved studies by electrospray mass spectrometry, *Anal. Chem. 75*, 6408–6414.
58. Wilson, D. J., Rafferty, S. P., and Konermann, L. (2005) Kinetic unfolding mechanism of the inducible nitric oxide synthase oxygenase domain determined by time-resolved electrospray mass spectrometry, *Biochemistry 44*, 2276–2283.
59. Simmons, D. A., Dunn, S. D., and Konermann, L. (2003) Conformational dynamics of partially denatured myoglobin studied by time-resolved electrospray mass spectrometry with online hydrogen-deuterium exchange, *Biochemistry 42*, 5896–5905.
60. Pan, J. X., Wilson, D. J., and Konermann, L. (2005) Pulsed hydrogen exchange and electrospray charge-state distribution as complementary probes of protein structure in kinetic experiments: Implications for ubiquitin folding, *Biochemistry 44*, 8627–8633.
61. Martin, S. E., Shabanowitz, J., Hunt, D. F., and Marto, J. A. (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem. 72*, 4266–4274.
62. Settlage, R. E., Christian, R. E., and Hunt, D. F. (2000) Method for structural characterization of biological moieties through HPLC separation, U.S. Patent. 6315905.
63. Kolakowski, B. M., Simmons, D. A., and Konermann, L. (2000) Stopped-flow electrospray ionization mass spectrometry: a new method for studying chemical reaction kinetics in solution, *Rapid Commun. Mass Spectrom. 14*, 772–776.
64. Kolakowski, B. M. and Konermann, L. (2001) From small-molecule reactions to protein folding: Studying biochemical kinetics by stopped-flow electrospray mass spectrometry, *Anal. Biochem. 292*, 107–114.
65. Bökenkamp, D., Desai, A., Yang, X., Tai, Y.-C., Marzluff, E. M., and Mayo, S. L. (1998) Microfabricated silicon mixers for submillisecond quench-flow analysis, *Anal. Chem. 70*, 232–236.
66. Oleschuk, R. D. and Harrison, D. J. (2000) Analytical microdevices for mass spectrometry, *Trends Anal. Chem. 19*, 379–388.
67. Creighton, T. E. (1992) *Protein Folding*, W.H. Freeman & Co., New York.
68. Creighton, T. E. (1984) Disulfide bond formation in proteins, *Methods Enzymol. 107*, 305–329.
69. Zhang, Y. H., Yan, X., Maier, C. S., Schimerlik, M. I., and Deinzer, M. L. (2002) Conformational analysis of intermediates involved in the *in vitro* folding pathways of recombinant human macrophage colony stimulating factor β by sulfhydryl group trapping and hydrogen/deuterium pulse labeling, *Biochemistry 41*, 15495–15504.
70. Apuy, J. L., Busenlehner, L. S., Russell, D. H., and Giedroc, D. P. (2004) Ratiometric pulsed alkylation mass spectrometry as a probe of thiolate reactivity in different metalloderivatives of Staphylococcus aureus pI258 CadC, *Biochemistry 43*, 3824–3834.
71. Apuy, J. L., Chen, X., Russell, D. H., Baldwin, T. O., and Giedroc, D. P. (2001) Ratiometric pulsed alkylation/mass spectrometry of the cysteine pairs in individual zinc fingers of MRE-binding transcription factor-1 (MTF-1) as a probe of zinc chelate stability, *Biochemistry 40*, 15164–15175.
72. Apuy, J. L., Park, Z. Y., Swartz, P. D., Dangott, L. J., Russell, D. H., and Baldwin, T. O. (2001) Pulsed-alkylation mass spectrometry for the study of protein folding and dynamics: development and application to the study of a folding/unfolding intermediate of bacterial luciferase, *Biochemistry 40*, 15153–15163.
73. Jha, S. K. and Udgaonkar, J. B. (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry, *J. Biol. Chem. 282*, 37479–37491.
74. Chen, H., Schuster, M. C., Sfyroera, G., Geisbrecht, B. V., and Lambris, J. D. (2008) Solution insights into the structure of the Efb/C3 complement inhibitory complex as revealed by lysine acetylation and mass spectrometry, *J. Am. Soc. Mass Spectrom. 19*, 55–65.
75. Mendoza, V. L. and Vachet, R. W. (2008) Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection, *Anal. Chem. 80*, 2895–2904.
76. Lim, J. and Vachet, R. W. (2003) Development of a methodology based on metal-catalyzed oxidation reactions and mass spectrometry to determine the metal binding sites in copper metalloproteins, *Anal. Chem. 75*, 1164–1172.
77. Gross, M. L., Gau, B. C., Sharp, J. S., and Rempel, D. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Anal. Chem. 81*, 6563–6571.
78. Hambly, D. M. and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *J. Am. Soc. Mass Spectrom. 16*, 2057–2063.
79. Gau, B. C., Chen, H., Zhang, Y., and Gross, M. L. (2010) Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins, *Anal. Chem. 82*, 7821–7827.
80. Chen, J. W., Rempel, D. L., and Gross, M. L. (2010) Temperature jump and fast photochemical oxidation probe submillisecond protein folding, *J. Am. Chem. Soc. 132*, 15502–15504.

81. Stocks, B. B., and Konermann, L. (2010) Time-Dependent Changes in side-chain solvent accessibility during cytochrome *c* folding probed by pulsed oxidative labeling and mass spectrometry, *J. Mol. Biol. 398*, 362–373.

82. Stocks, B. B. and Konermann, L. (2009) Structural characterization of short-lived protein unfolding intermediates by laser-induced oxidative labeling and mass spectrometry, *Anal. Chem. 81*, 20–27.

83. Stocks, B. B., Rezvanpour, A., Shaw, G. S., and Konermann, L. (2011) Temporal development of protein structure during S100A11 folding and dimerization probed by oxidative labeling and mass spectrometry, *J. Mol. Biol. 409*, 669–679.

84. Pan, Y., Brown, L., and Konermann, L. (2011) Kinetic folding mechanism of an integral membrane protein examined by pulsed oxidative labeling and mass spectrometry, *J. Mol. Biol. 410*, 146–158.

85. Konermann, L., Stocks, B. B., Pan, Y., and Tong, X. (2010) Mass spectrometry combined with oxidative labeling for exploring protein structure and folding, *Mass Spectrom. Rev. 29*, 651–667.

86. Kiselar, J. G. and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting, *J. Mass Spectrom: JMS 45*, 1373–1382.

87. Konermann, L., Pan, Y., and Stocks, B. B. (2011) Protein folding mechanisms studied by pulsed oxidative labeling and mass spectrometry, *Curr. Opin. Struct. Biol. 21*, 634–640.

88. Xu, G. H. and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem Rev 107*, 3514–3543.

89. Fändrich, M., Tito, M. A., Leroux, M. R., Rostom, A. A., Hartl, F. U., Dobson, C. M., and Robinson, C. V. (2000) Observation of the noncovalent assembly and disassembly pathways of the chaperone complex MtGimC by mass spectrometry, *Proc. Natl. Acad. Sci. USA 97*, 14151–14155.

90. Levy, E. D., Boeri Erba, E., Robinson, C. V., and Teichmann, S. A. (2008) Assembly reflects evolution of protein complexes, *Nature (London) 453*, 1262–1265.

91. Sharon, M., Witt, S., Glasmacher, E., Baumeister, W., and Robinson, C. V. (2007) Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome, *J. Biolo. Chem. 282*, 18448–18457.

92. Sobott, F., Benesch, J. L. P., Vierling, E., and Robinson, C. V. (2002) Subunit exchange of multimeric protein complexes. Real time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry, *J. Biol. Chem. 277*, 38921–38929.

93. Painter, A. J., Jaya, N., Basha, E., Vierling, E., Robinson, C. V., and Benesch, J. L. (2008) Real-time monitoring of protein complexes reveals their quaternary organization and dynamics, *Chem. Biol. 15*, 246–253.

94. Stengel, F., Baldwin, A. J., Painter, A. J., Jaya, N., Basha, E., Kay, L. E., Vierling, E., Robinson, C. V., and Benesch, J. L. (2010) Quaternary dynamics and plasticity underlie small heat shock protein chaperone function, *Proc. Natl. Acad. Sci. USA 107*, 2007–2012.

95. Eyles, S. J. and Gierasch, L. M. (2010) Nature's molecular sponges: small heat shock proteins grow into their chaperone roles, *Proc. Natl. Acad. Sci. USA 107*, 2727–2728.

96. Chevreux, G., Atmanene, C., Lopez, P., Ouazzani, J., Van Dorsselaer, A., Badet, B., Badet-Denisot, M. A., and Sanglier-Cianferani, S. (2011) Monitoring the dynamics of monomer exchange using electrospray mass spectrometry: the case of the dimeric glucosamine-6-phosphate synthase, *J. Am. Soc. Mass Spectrom. 22*, 431–439.

97. Van Pelt, C. K., Zhang, S., and Henion, J. D. (2002) Characterization of a fully automated nanoelectrospray system with mass spectrometric detection for proteomic analyses, *J. Bio. Tech.: JBT 13*, 72–84.

98. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J. E., Ruotolo, B. T., Robinson, C. V., and Bowers, M. T. (2009) Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease, *Nat. Chem. 1*, 326–331.

99. Smith, D. P., Radford, S. E., and Ashcroft, A. E. (2010) Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry, *Proc. Natl. Acad. Sci. USA 107*, 6794–6798.

100. Michaelis, L. and Menten, M. L. (1913) Die kinetik der invertinwirkung, *Biochem. Z. 49*, 333–339.

101. Huggett, A. S. G. and Nixon, D. A. (1957) Enzymic determination of blood glucose, *Biochem. J. 66*, 12P.

102. Wallenfels, K. (1962) Beta-galactosidase (crystalline), *Methods Enzymol. 5*, 212–219.

103. Northrop, D. B. and Simpson, F. B. (1997) Beyond enzyme kinetics: Direct determination of mechanisms by stopped-flow mass spectrometry, *Bioorg. Med. Chem. Lett. 5*, 641–644.

104. Smith, L. A. and Caprioli, R. M. (1983) Following enzyme catalysis in real time inside a fast atom bombardment mass spectrometer, *Biomed. Mass Spectrom. 10*, 98–102.

105. Smith, L. A. and Caprioli, R. M. (1984) Enzyme reaction rates determined by fast atom bombardment mass-spectrometry, *Biomed. Mass Spectrom. 11*, 392–395.

106. Hsieh, F. Y. L., Tong, X., Wachs, T., Ganem, B., and Henion, J. (1995) Kinetic monitoring of enzymatic reactions in real time by quantitative high performance liquid chromatography mass spectrometry, *Anal. Biochem. 229*, 20–25.

107. Anderson, K. S., Sikorski, J. A., and Johnson, K. A. (1988) A tetrahedral intermediate in the EPSP synthase reaction observed by rapid quench kinetics, *Biochemistry 27*, 7395–7406.

108. Anderson, K. S., Sikorski, J. A., and Johnson, K. A. (1988) Evaluation of 5-enolpyruvoylshikimate-3-phosphate synthase substrate and inhibitor binding by stopped-flow and equilibrium fluorescence measurements, *Biochemistry 27*, 1604–1610.

109. Brown, E. D., Marquardt, J. L., Lee, J. P., Walsh, C. T., and Anderson, K. S. (1994) Detection and characterization of a phospholactoyl-enzyme adduct in the reaction catalyzed by UDP-*N*-acetylglucosamine enolpyruvoyl transferase, MurZ, *Biochemistry 33*, 10638–10645.

110. Paiva, A. A., Tilton, R. F., Crooks, G. P., Huang, L. Q., and Anderson, K. S. (1997) Detection and identification of transient enzyme intermediates using rapid mixing, pulsed-flow electrospray mass spectrometry, *Biochemistry 36*, 15472–15476.

111. Li, Z., Sau, A. K., Shen, S., Whitehouse, C., Baasov, T., and Anderson, K. S. (2003) A snapshot of enzyme catalysis using electrospray ionization mass spectrometry, *J. Am. Chem. Soc. 125*, 9938–9939.

112. Zechel, D. L., Konermann, L., Withers, S. G., and Douglas, D. J. (1998) Pre-steady state kinetic analysis of an enzymatic reaction monitored by time-resolved electrospray ionization mass spectrometry, *Biochemistry 37*, 7664–7669.

113. Wilson, D. J. and Konermann, L. (2004) Mechanistic studies on enzymatic reactions by electrospray ionization MS using a capillary mixer with adjustable reaction chamber volume for time-resolved measurements, *Anal. Chem. 76*, 2537–2543.

114. Brown, R. P. A., Aplin, R. T., and Schofield, C. T. (1996) Inhibition of TEM-2,β-lactamase from Escherichia coli by clavulanic acid: observation of intermediates by electrospray ionization mass spectrometry, *Biochemistry 35*, 12421–12432.

115. Gao, H., Yu, Y., and Leary, J. A. (2005) Mechanism and kinetics of metalloenzyme phosphomannose isomerase: measurement of dissociation constants and effect of zinc binding using ESI–FTICR mass spectrometry, *Anal. Chem. 77*, 5596–5603.

116. Pi, N., Hoang, M. B., Gao, H., Mougous, J. D., Bertozzi, C. R., and Leary, J. A. (2005) Kinetic measurements and mechanism determination of Stf0 sulfotransferase using mass spectrometry, *Anal. Biochem. 341*, 94–104.

117. Pi, N., Yu, Y., Mougous, J. D., and Leary, J. A. (2004) Observation of a hybrid random ping-pong mechanism of catalysis for NodST: a mass spectrometry approach, *Protein Sci.: a Publication Protein Soc. 13*, 903–912.

118. Yu, Y., Kirkup, C. E., Pi, N., and Leary, J. A. (2004) Characterization of noncovalent protein-ligand complexes and associated enzyme intermediates of GlcNAc-6-*O*-sulfotransferase by electrospray ionization FT-ICR mass spectrometry, *J. Am. Soc. Mass Spectrom. 15*, 1400–1407.

119. Pi, N., Meyers, C. L., Pacholec, M., Walsh, C. T., and Leary, J. A. (2004) Mass spectrometric characterization of a three-enzyme tandem reaction for assembly and modification of the novobiocin skeleton, *Proc. Natl. Acad. Sci. USA 101*, 10036–10041.

120. Houston, C. T., Taylor, W. P., Widlanski, T. S., and Reilly, J. P. (2000) Investigation of enzyme kinetics using quench-flow techniques with MALDI-TOF mass spectrometry, *Anal. Chem. 72*, 3311–3319.

121. Nichols, K. P. and Gardeniers, H. J. (2007) A digital microfluidic system for the investigation of pre-steady-state enzyme kinetics using rapid quenching with MALDI-TOF mass spectrometry, *Anal. Chem. 79*, 8699–8704.

122. Wei, J., Buriak, J. M., and Siuzdak, G. (1999) Desorption-ionization mass spectrometry on porous silicon, *Nature (London) 399*, 243–246.

123. Nichols, K. P., Azoz, S., and Gardeniers, H. J. (2008) Enzyme kinetics by directly imaging a porous silicon microfluidic reactor using desorption/ionization on silicon mass spectrometry, *Anal. Chem. 80*, 8314–8319.

124. Ozbal, C. C., LaMarr, W. A., Linton, J. R., Green, D. F., Katz, A., Morrison, T. B., and Brenan, C. J. (2004) High throughput screening via mass spectrometry: a case study using acetylcholinesterase, *Assay Drug Dev. Technol. 2*, 373–381.

125. Greis, K. D. (2007) Mass spectrometry for enzyme assays and inhibitor screening: an emerging application in pharmaceutical research, *Mass Spectrom. Rev. 26*, 324–339.

7

PROTEIN INTERACTIONS: A CLOSER LOOK AT THE STRUCTURE–DYNAMICS–FUNCTION TRIAD

In Chapters 4–6, we discussed various mass spectrometry (MS)-based techniques to study biomolecular higher order structure and conformational dynamics under a variety of conditions. Characterization of protein structure and dynamics, including those of non-native states, is often key to understanding how these macromolecules carry out their "duties" via a sophisticated web of interactions with other proteins, polysaccharides, nucleic acids, and small molecule ligands. This chapter gives an overview of some of the vast arsenal of experimental techniques that can be used to extract both qualitative and quantitative information on protein–ligand interaction. The emphasis will be made on protein interactions with their physiological partners and various therapeutic molecules for which they are targets. We will first consider several approaches to monitor protein–small ligand interactions and measure their binding energy. Particular attention will be paid to complexes that are unstable in the gas phase. We will then proceed to discuss other methods that are used to obtain quantitative information on protein–protein interactions. Finally, we will consider several examples where the ability to characterize protein conformation and dynamics is critical to understanding their functional properties.

7.1. DIRECT METHODS OF MONITORING PROTEIN INTERACTIONS WITH THEIR PHYSIOLOGICAL PARTNERS IN SOLUTION BY ESI MS: FROM SMALL LIGANDS TO OTHER BIOPOLYMERS

Although the earliest applications of electrospray ionization (ESI) MS focused on measuring mass and obtaining sequence information of single-chain biopolymers, the soft nature of this ionization technique prompted several researchers to explore the possibility of maintaining the structural integrity of protein–small ligand complexes throughout the desorption process. In Section 4.1 we saw that the gentle nature of the ESI process allows the quaternary structure of proteins to be preserved in many cases, but its ability to preserve other types of noncovalent complexes in the gas phase also means that many other associations may be probed with this techniques. As early as 1991, Ganem, Li, and Henion (1) observed intact complexes formed by a small immunosuppressive binding protein and small immunosuppressive agents (e.g., rapamycin and its analogue). This report was soon followed by two other important studies in which intact non-covalent complexes of lysozyme and its natural hexasaccharide substrate (2), as

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

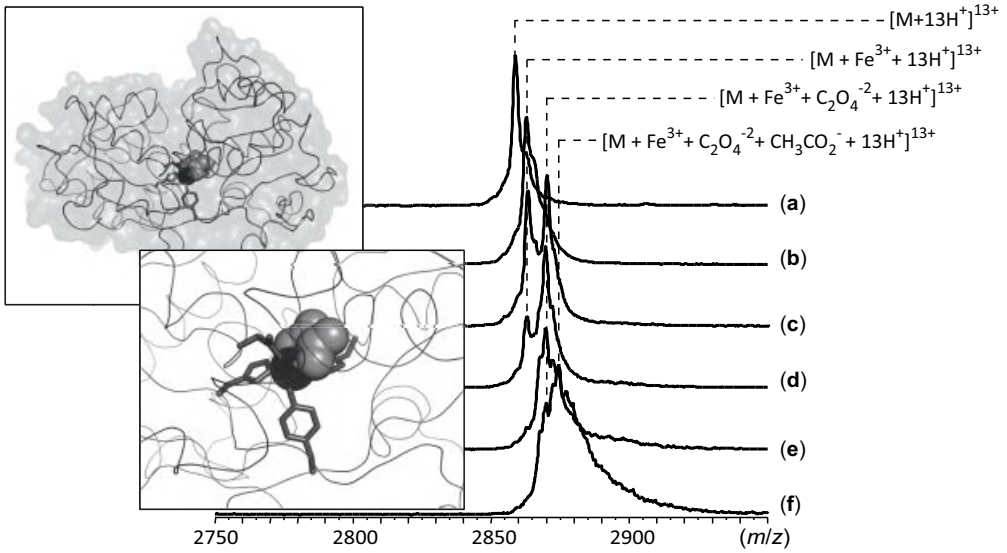


Figure 7.1. Dissociation of the ternary complex [*h*Tf/2N·Fe^{3+}·$C_2O_4^{2-}$] in the gas phase: ion peaks of a 13+ charge state for control apo-*h*Tf/2N in $H_2O/CH_3OH/CH_3CO_2H$ (*a*) and holo-*h*Tf/2N in 10 mM $NH_4CH_3CO_2$ (*b*–*f*). The efficiency of ion collisional activation decreases from (b) to (f) following monotonic decrease of the skimmer potential and temperature in the ESI interface. The insets show the crystal structure of *h*Tf/2N serum transferrin (PDB accession number 1A8E) and the expanded view of the iron-binding site (iron is shown as a black sphere, and the synergistic anion is represented with a gray space-filled model; the four side chains coordinating iron are represented with stick models). [Adapted with permission from (10). Copyright © 2001 American Chemical Society.]

well as of the myoglobin–heme group complex (3), were observed. The list of ligands that bind to proteins strongly enough to survive the transition from solution to the gas phase has been expanded dramatically in the following couple of years to include metal ions (4,5), nucleotides (6), oligonucleotides (7,8), as well as peptides and proteins (9).

The basic principles of such measurements are very close to those discussed in Section 4.1 when we considered applications of native ESI MS to characterization of protein quaternary structure, although there are several important considerations that are specific to monitoring protein–small ligand binding. First, such measurements typically require better mass resolution and accuracy, especially when a small-molecule ligand (e.g., a metal ion) is bound to a relatively large protein. Second, in many cases multiple ligands may be bound to the same protein, and their cumulative mass alone may not be very informative. Finally, protein–small ligand complexes are frequently more prone to dissociation in the gas phase compared to protein–protein associations due to the lower number of bonds that keep them together in solution and that can survive transition to the gas phase. Ironically, the experimental conditions that may alleviate this latter problem (very gentle conditions in the ESI interface) almost invariably result in poor protein ion desolvation, making it difficult to attain high mass accuracy in such measurements.

Nevertheless, in many cases it is possible to design an experimental scheme, where all of the problems listed in the previous paragraph can be successfully addressed, as illustrated in Figure 7.1 with a 37 kDa metalloprotein (N-lobe of human serum transferrin, *h*Tf/2N) as an example. This protein binds a range of metal ions, most notably ferric ion (Fe^{3+}) together with a synergistic anion (e.g., carbonate), which participates in coordinating the metal ion. Although ESI MS measurements produce a single mass readout, corresponding to the combined mass of the protein, metal ion, and the synergistic anion, manipulation of the complex ion in the gas phase allows the masses of both ligands to be measured. This is done by increasing the internal energy of the complex ion in the gas phase in a stepwise fashion via collisional activation in the ESI MS interface, which leads to partial dissociation of the complex. While the protein–metal binding is very strong and survives significant collisional excitation, dissociation of the synergistic anion is observed even following relatively modest excitation (Fig. 7.1*a*). In contrast, very gentle conditions in the ESI MS interface result in inefficient ion desolvation, leading to formation of nonspecific adducts (e.g., attachment of acetate to the ternary complex in Fig. 7.1*f*). Therefore, confident identification of both small ligands cannot be made in a single measurement, but requires a set of mass spectra to be acquired under various conditions in the ESI MS interface.

High mass accuracy is another important requirement in such experiments, and the measurements presented in Figure 7.1 provided mass accuracy better than 0.01%. Note

that these high-accuracy measurements yielded a somewhat unexpected result: it was found that the mass of the synergistic anion was 18 u higher than that expected for carbonate binding (10). Instead, the mass matches that of oxalate, a ubiquitous dianion. It seemed strange at the time this work was carried out that oxalate, rather than carbonate, should serve as a synergistic anion (the latter was present in the protein solution in large excess, since all protein samples were stored in ammonium bicarbonate solution, while their exposure to oxalate was possible only during the expression step, where oxalate was a minor component of culture media). However, later studies revealed that oxalate was bound much more strongly compared to carbonate (11), providing an explanation as to why it remained bound to *h*Tf/2N, despite the presence of excess carbonate in the protein solution.

While the work discussed above is a very nice illustration of the power of MS in elucidating the nature of protein-binding partners, remember that this level of detail might be difficult, or indeed impossible, to achieve in many cases. This point is illustrated in Figure 7.2, which shows the results of ESI MS investigation of ibuprofen binding to a 66 kDa BSA, which acts as a scavenger in plasma and binds a range of metals and small organic molecules while in circulation. Although comparison of the mass spectra of albumin acquired in the absence and in the presence of ibuprofen in solution reveals an obvious mass shift, the magnitude of this mass increase (∼300 Da) is not very informative, as it does not match the mass of ibuprofen (206 Da). The situation is further exacerbated by the convoluted peak shapes of albumin ions, which probably reflects both its scavenger function and the fact that it is a common target of nonenzymatic post-translational modifications (e.g., glycation) (12). Nevertheless, even in this case, careful analysis of the protein ion peak shapes reveals mass increases of 206 and 412 Da (indicated with arrows in Fig. 7.2), which are indicative of ibuprofen binding to albumin in 1:1 and 2:1 stoichiometry, consistent with the presence of two binding sites on the protein surface that can accommodate this drug (13).

Finally, note that failure to observe a protein–small ligand complex in the gas phase does not necessarily indicate its absence in solution, particularly if the interaction is primarily hydrophobic. We already saw in Section 4.1 that gas-phase dissociation of macromolecular assemblies is frequently encountered in the studies of quaternary structure of proteins using native ESI MS, although these processes can be easily identified in most cases, due to very unusual multiple charging patterns of the fragmentation products. It may seem counterintuitive that gas-phase dissociation of protein–small ligand complexes presents a more significant challenge vis-à-vis assigning the origin of subcomplex ions, when compared to the more substantial polypeptide assemblies already considered. If a departing ligand does not possess sufficient physical dimensions to accommodate multiple charges, its dissociation from the host protein would not result in asymmetric charge partitioning commonly observed in dissociation of protein assemblies in ESI

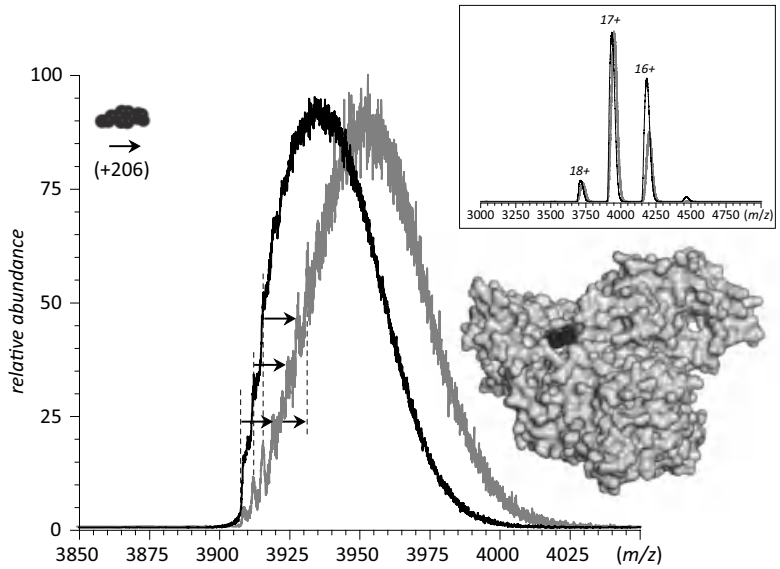


Figure 7.2. Binding of ibuprofen to serum albumin revealed by ESI MS under near-native conditions. The black and gray traces represent mass spectra of bovine serum albumin (BSA) alone and with significant molar excess of ibuprofen, respectively. The crystal structure of an ibuprofen–albumin complex shows two drug molecules bound to the protein (the second one is hidden from the view, as its binding site is located on the opposite side of the protein). [Data courtesy of Dr. Agya Frimpong (UMass-Amherst).]

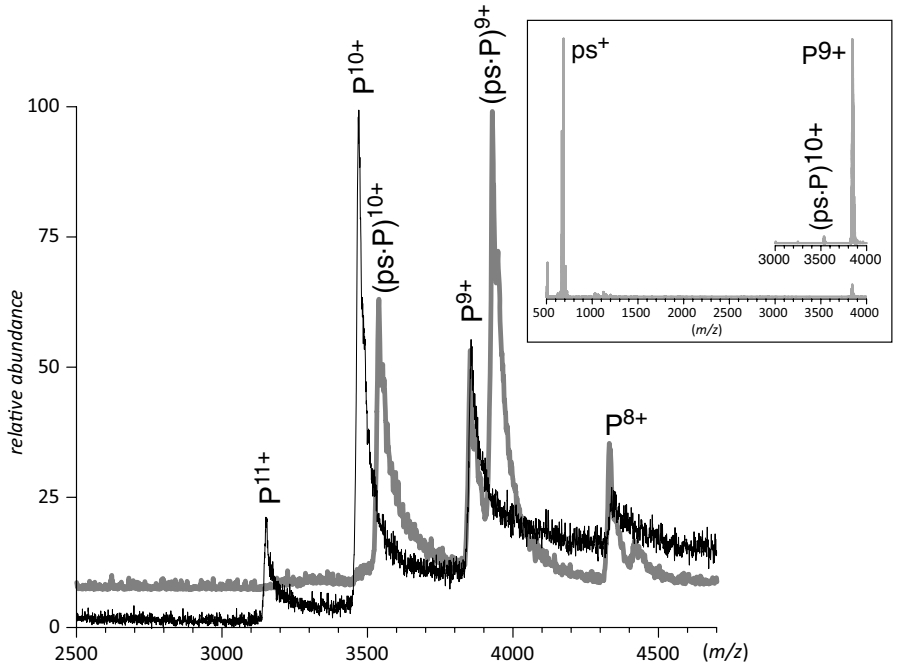


Figure 7.3. The ESI mass spectrum of pepsin (P) acquired at pH 2.3 (black trace). The gray trace shows a mass spectrum of pepsin incubated with excess of pepstatin (ps). Inset: CAD mass spectrum of $(ps{\cdot}P)^{10+}$ ion suggests that enzyme–inhibitor complex dissociation in the gas phase proceeds via charge separation. [Adapted with permission from (14)].

MS. Therefore, the resulting gas-phase fragment ions might not be easily distinguishable from those generated in solution. However, as long as at least one charge can be placed on a ligand, its dissociation from the protein in the gas phase will affect the charge-state distribution of the latter, providing a means to identify gas-phase dissociation products. This is illustrated in Figure 7.3, which shows the ESI mass spectrum of a mixture of pepsin with the peptide inhibitor pepstatin, acquired under conditions when formation of the enzyme–inhibitor complex in solution is highly favorable. Although the peaks arising from the apoprotein are prominent in this spectrum, a close examination of the charge-state distribution of pepsin ions reveals a shift to a lower number of charges when compared to the mass spectrum of pepsin acquired in the absence of pepstatin. This apparent charge reduction clearly indicates that the ligand-free pepsin ions in the ESI mass spectrum of the pepsin–pepstatin mixture are produced upon dissociation of the enzyme–inhibitor complex in the gas phase, which proceeds via energetically favorable charge separation.

Although reports of failure to observe noncovalent complexes are unsurprisingly scarce in the current literature, it remains a distinct possibility in practice, and researchers need to be aware of it. Nevertheless, the ability of ESI MS to detect protein interactions with a wide variety of partners in solution is now indisputable, and this technique is now rightly considered a routine tool for screening protein–ligand interactions (15,16).

7.2. ASSESSMENT OF BINDING AFFINITY WITH DIRECT ESI MS APPROACHES

The success of the early studies aimed at probing protein interactions with their physiological partners provided strong motivation for comprehensive exploration of the utility of ESI MS for characterization of such interactions in solution. Several groups evaluated the possibility to correlate binding properties in solution and complex ion stability in the gas phase for a variety of systems (17–19). Although surprisingly good correlation was found to exist between the solution- and gas-phase binding properties in select cases (19), significant deviations were observed for many other systems (20,21). These observations of frequent deviation of binding characteristics in solution from those in the gas phase indicated that it might be very dangerous to simply equate the solution- and gas-phase properties of noncovalent complexes. As noted by Ganem, Henion and co-workers (22), "noncovalent binding of macromolecules usually involves a broad range of weak interactions, ranging from hydrogen bonding and electrostatic effects to ... π-stacking and hydrophobic effects [with] each type of interaction responding differently to changes in solvation, as would occur during ionspray or electrospray ionization".

Apart from correlating the energetics of gas-phase noncovalent complexes with binding energies in solution, multiple attempts had been made to determine K_D values for protein–ligand complexes in solution using intensities of

ionic signals in ESI MS to calculate concentrations of complexes and unbound species in solution. For example, Ganem and co-workers (22) demonstrated that direct monitoring of the relative abundance of complex ions formed by a model glycopeptide and antibiotics from the vancomycin family provided a correct order of magnitude for the peptide–drug association constants. Similar conclusions have been reached in a number of other studies, where MS was shown to be capable of providing correct quantitative assessment of binding affinity in solution (23–26). Although this approach continued to be used to evaluate the strength of protein–ligand binding in solution, its limitations soon became apparent (27). There are multiple reasons why the reports of measurements using this approach are frequently met with at least some skepticism. First, as we already discussed in the beginning of this chapter, there is always a distinct possibility of the protein–ligand complex dissociating in the gas phase. This would obviously modulate the ionic signal for all players, making it reflect not only the solution-phase equilibria, but also stability of the protein–ligand complex in the gas phase. Second, even though in many practical situations (i.e., if the protein concentration is $<10\,\mu$M) the abundance of the ionic signal is proportional to the species concentration in solution (28), the response factors vary across the analyte pool and, in fact, depend on solution composition (due to ion suppression). Obviously, this variation needs to be taken into account when relating analyte signal intensity in ESI MS to its concentration in solution (29–31). Finally, one must be aware of the fact that the ESI process itself may artificially enhance the apparent concentration of complexes by ramping up the analyte concentration in the ESI-generated droplets prior to ion generation (32).

An alternative approach to obtaining quantitative binding information from direct ESI MS measurements exploits the ability of this technique to monitor formation of multiple protein–ligand complexes simultaneously when a range of binding candidates are present in solution. Gao and co-workers (33) used ESI MS to determine relative binding affinities of a protein to a variety of ligands by measuring the abundance of respective complex ions in mass spectra generated from a solution containing the protein of interest and a library of ligands (33,34). This formed the basis of the "competitive binding" measurement of ligand binding by ESI MS. Modifications of this method extended the scope of its applications to protein–protein interactions, providing very high precision in K_D measurements (35), and allowing, in some cases, relatively low affinities to be evaluated, provided there is a set of reference binders with known affinities (36). However, problems frequently arise when the competitive binding approach is applied to a large heterogeneous protein and a set of relatively small binders. The resolution in this case may not be sufficient to allow various complexes to be distinguished from one another. For example, looking at Figure 7.2, one could easily see that this problem would likely be encountered if this method were used to rank albumin affinities with a set of small organic molecules the size of ibuprofen. An elegant solution to this problem was offered by Zenobi and co-workers (37), who ranked the binding affinities within a set of small molecules by monitoring the signal intensity changes for the unbound species, that is by following free ligand depletion, rather than complex formation. More detailed discussion of various approaches utilizing direct ESI MS measurements to quantify the strength of protein–ligand complexes in solution can be found in a recent review by Hannah et al. (16).

7.3. INDIRECT CHARACTERIZATION OF NON-COVALENT INTERACTIONS UNDER PHYSIOLOGICAL AND NEAR-PHYSIOLOGICAL CONDITIONS

Hydrophobic interactions are often critical contributors to protein–ligand interactions in solution but, as we have already seen, they frequently do not survive transition to the gas phase. Should this lead to dissociation of even a fraction of the protein–ligand complexes in the gas phase, the consequences would be dire for the affinity measurements approaches using direct ESI MS discussed in Section 7.2. This section presents several methods that were designed to overcome this problem by probing noncovalent interactions indirectly, that is to say, without relying on survival of the protein–ligand complexes upon their transition from solution to the gas phase.

7.3.1. Assessment of Ligand Binding by Monitoring Dynamics of "Native" Proteins with Hydrogen–deuterium Exchange (HDX MS)

One interesting example of such an indirect approach to evaluate protein–ligand binding properties in solution comes from our own work with cellular retinoic acid binding protein I (CRABP I), a transporter of all-trans retinoic acid (RA), which is a major physiologically active metabolite of vitamin A. The X-ray crystal structure suggests that the RA–CRABP I complex is stabilized primarily through a network of hydrophobic interactions, although there is also an intermolecular salt bridge that makes a significant contribution toward stability of this complex (K_D in the subnanomolar range). In addition to RA, CRABP I also binds (albeit with significantly lower affinity) the 9-cis and 13-cis stereoisomers of RA, which are considered to be noncognate ligands for this protein. Although the ESI mass spectra of the CRABP I–RA mixture acquired under mild conditions in the ESI interface show prominent signals corresponding to protein–ligand complex ions (Fig. 7.4), their relative intensities do not agree with ranking of binding affinities in solution (38).

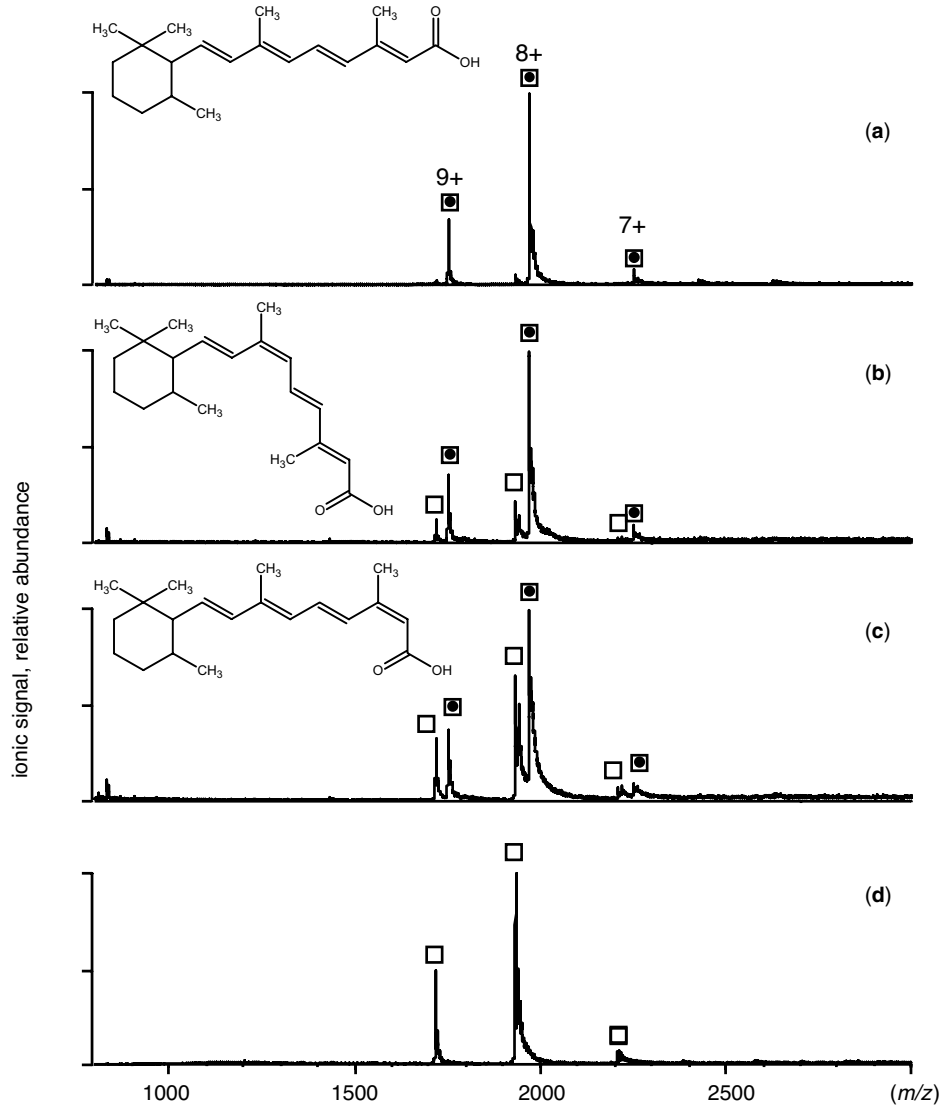


Figure 7.4. The ESI mass spectra of CRABP I mixed with RA isomers and analogues all-trans, (*a*) 13-cis, (*b*) 9-cis, (*c*) no ligand added (*d*). The protein/ligand ratio (where appropriate) is 1:5. All spectra are acquired under "near-native" solution conditions. The ESI interface settings were adjusted to minimize collisional activation/dissociation. [Adapted with permission from (38). Copyright © 2003 American Society for Mass Spectrometry.]

As already discussed in Chapter 1, it is now realized that in many cases protein folding and function should not be considered separately from each other. This notion is often used when studying protein–protein interactions (39,40), however, it can also be very useful when considering protein interaction with small organic ligands. A common assumption here is that binding reduces flexibility, which has been proven experimentally in the case of several retinoid carriers (41–43). Therefore, changes in protein backbone dynamics upon ligand introduction can be used as indicators of the protein–ligand interaction. This hypothesis was tested by measuring the HDX kinetics of CRABP I under near-native conditions (Fig. 7.5). Under these conditions, HDX clearly follows the EX2 regime (detailed discussion of HDX experiments are presented in Chapter 5), and the overall kinetics exhibits three phases (fast, intermediate, and slow). As we discussed (Section 5.3), the slow exchange phase usually corresponds to global unfolding events that occur very rarely under native conditions. According to Eq. 5-3-4, the apparent rate constant corresponding to this phase would be related to both intrinsic exchange rate k_{int} and the equilibrium constant for the unfolding process K_{unfold}:

$$k^{\text{HDX}} = k_{\text{int}} K_{\text{unfold}} \qquad (7\text{-}3\text{-}1)$$

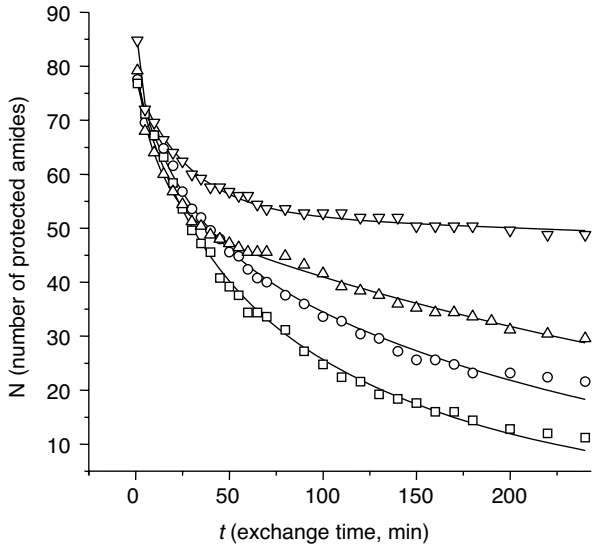


Figure 7.5. The HDX kinetics of CRABP I at 36°C under "near-native" conditions in the absence of ligands (squares) and in the presence of RA isomers (13-cis, circles; 9-cis, triangles; and all-trans, inverted triangles). [Adapted with permission from (38). Copyright © 2003 American Society for Mass Spectrometry.]

Since the equilibrium constant K_{unfold} relates to the free energy of the native conformation in the following fashion:

$$\Delta G_{\text{N}} = -RT\ln(K_{\text{unfold}}) \qquad (7\text{-}3\text{-}2)$$

and the protein–ligand binding energy can be estimated as the free energy change of the native conformation induced by the ligand (the asterisk indicates presence of the ligand in the system):

$$\Delta G_{\text{binding}} = \Delta G_{\mathbf{N}^*} - \Delta G_{\mathbf{N}} = \Delta\Delta G \qquad (7\text{-}3\text{-}3)$$

one can estimate the binding energy simply by comparing the apparent HDX rate constant measured in the presence and in the absence of the ligand:

$$\Delta G_{\text{binding}} = RT\ln(k_{\text{HDX}}/k^*_{\text{HDX}}) \qquad (7\text{-}3\text{-}4)$$

An implicit assumption used here is that there is no significant conformational difference between the apo-protein and its holo-form, so that ligand-induced structure stabilization would be the only factor contributing to the free energy change upon ligand binding by the protein (Eq. 7-3-3). In other words, this approach can be safely used only if the protein–ligand binding process conforms to the "lock-and-key" mechanism (although there may be significant conformational changes during the various stages of the binding process, as long as the end-points of the binding reaction are structurally similar, the thermodynamic treatment presented here can be safely used). Using this approach to evaluate CRABP I affinity to various retinoids not only provides correct ranking of RA stereoisomers, but also gives an estimation of the protein–RA binding energy, which is reasonably close to that obtained by calorimetric measurements (38).

A discussion of this method would be incomplete without mentioning that the assumption that "binding reduces flexibility" is not universally true. It is often anticipated that specific, high-affinity binding proceeds through formation of well-defined interactions between a protein and its ligand, leading to a loss of conformational entropy, which would need to be more than offset by favorable enthalpic interactions so that binding would be thermodynamically favored. However, an increase of the conformational entropy of the residues interacting with the ligand may also promote the binding process, as positive ΔS would lead to more negative (and, therefore, more favorable) ΔG, even if the enthalpic contribution ΔH is rather insignificant (in this case $\Delta G = \Delta H - \text{T}\Delta S \approx -T\Delta S$) (44). Indeed, there are several documented examples where binding actually leads to a detectable increase in protein flexibility (45–47). These findings, however, are likely to have little bearing on the validity of the method discussed in this section. Indeed, the observed increases of protein flexibility upon ligand binding refer to local events, typically at or near the binding site(s) (44). Such behavior would manifest itself as an increase of the HDX rates in the fast (local fluctuations) and medium (partially unstructured intermediate states) phases of the overall HDX pattern. None of these measurements are used to determine the binding energy, whose calculation (Eq. 7-3-4) is based solely on the parameters of the slow phase (rare global unfolding events).

7.3.2. PLIMSTEX and Related Techniques: Binding Assessment by Monitoring Conformational Changes with HDX MS in Titration Experiments

A powerful method that uses HDX to quantify protein–ligand interaction (PLIMSTEX, short for Protein–Ligand Interactions in solution by MS, Titration, and H/D EXchange) was developed by Gross and co-workers (48,49). PLIMSTEX is based upon the assumption that ligand binding results in a change in HDX rates (e.g., due to different solvent exposure of the apo- and holo-forms of the protein). The method is analogous to titration monitored by spectroscopic methods. The approach, although indirect, overcomes a major difficulty of ESI MS to measure, without discrimination, solution concentrations at any point, including at equilibrium. In addition to binding energetics, this method is capable of determining protein–ligand complex stoichiometry.

The protein is first equilibrated with different concentrations of ligand in aqueous buffer, and HDX reactions are then initiated by diluting these protein solutions in D_2O without changing any other solution parameters (salt content, buffer

compositions, etc.). After reaching a near-steady state (after at least 1 h), the exchange is quenched, followed by quick desalting under slow exchange conditions and mass analysis. Since the slow exchange conditions typically cause the ligand(s) to dissociate, mass measurements simply reveal the number of deuterium atoms on solvent-accessible amides. A plot of these numbers [deuterium uptake (ΔD) as a function of the total ligand concentration ([Lig_T]] gives the PLIMSTEX curve. The data analysis step seeks to find the best fit to the PLIMSTEX curve, using as fitting parameters the cumulative binding constants $\beta_i = K_1 \cdot K_2 \cdot \cdots \cdot K_i$, where K_j is a microscopic binding constant for the jth ligand, and the respective deuterium shifts D_0 and ΔD_i, where D_0 represents the deuterium uptake of the apo-form of the protein:

$$\Delta D(\beta_1, \ldots, \beta_n, D_0, D_1, \ldots, D_n, [\mathrm{Lig}_T]) = D_0 - \sum_{i=1}^{n} \Delta D_1 \frac{[\mathrm{Prot} \cdot \mathrm{Lig}_i]}{[\mathrm{Prot}_T]} \qquad (7\text{-}3\text{-}5)$$

A positive ΔD_i indicates that binding of the ith ligand to the protein leads to more protection (less solvent exposure and/or conformational flexibility) compared to the apo-form. A negative ΔD_i points to the formation of a more open (or more flexible) structure relative to the ligand-free form of the protein.

Each form of the protein–ligand complex Prot·Lig_i in Eq. 7-3-5 is expected to generate a deuterium shift, which is the difference between the deuteration level of this complex and the apo-form of the protein. While [Prot_T] is known, the binding fraction for each species in Eq. 7-3-5 is calculated for each sequential ligand attachment:

$$\frac{[\mathrm{Prot} \cdot \mathrm{Lig}_i]}{[\mathrm{Prot}_T]} = \frac{\beta_i [\mathrm{Lig}]^i}{1 + \sum_{i=1}^{n} \beta_i [\mathrm{Lig}]^i} \qquad (7\text{-}3\text{-}6)$$

Calculation of the binding fractions may not be trivial, since the concentration of free ligand in solution may not be known. However, its value can be inferred from the total ligand concentration using an ordinary differential equation, which is an expression of the first-order derivative of [Lig] with respect to [Lig_T]:

$$\frac{d}{d[\mathrm{Lig}_T]}[\mathrm{Lig}] = \frac{\sum_{i=0}^{n} \beta_i \cdot [Lig]^i}{\sum_{i=1}^{n} (i+1)\, \beta_i [\mathrm{Lig}]^i + [\mathrm{Prot}_T] \cdot \left(\sum_{i=1}^{n} i^2 \cdot \beta_i \cdot [\mathrm{Lig}]^{i-1} \right) - [\mathrm{Lig}_T] \left(\sum_{i=1}^{n} i\, \beta_i [\mathrm{Lig}]^{i-1} \right)} \qquad (7\text{-}3\text{-}7)$$

In the case of 1 : 1 binding stoichiometry, Eq. 7-3-7 can be rewritten in a much less intimidating form:

$$\frac{d}{d[\mathrm{Lig}_T]}[\mathrm{Lig}] = \frac{1 + K \cdot [\mathrm{Lig}]}{1 + 2K\,[\mathrm{Lig}] + K\,[\mathrm{Prot}_T] - K\,[\mathrm{Lig}_T]} \qquad (7\text{-}3\text{-}8)$$

where K is the protein–ligand binding constant. Integration of Eq. 7-3-7 can be carried out numerically using commercially available packages, such as MathCAD (49). Once this task is accomplished, the remaining sets of unknowns (β_i and ΔD_i) are determined in a sequence of trials.

The ability of PLIMSTEX to determine both binding stoichiometry and affinity associated with a wide range of protein–ligand interactions is particularly useful for quantifying protein–metal ion interactions (50,51); Figure 7.6 shows an application of PLIMSTEX to probe Ca^{2+} binding to porcine calmodulin. More recently, this technique has also been applied to characterize protein–DNA interactions (52) and antibody–antigen association (53). Adaptation of PLIMSTEX to studies of protein self-association (54) has been termed SIMSTEX (Self-association Interactions using Mass Spectrometry, Self-Titration, and H/D

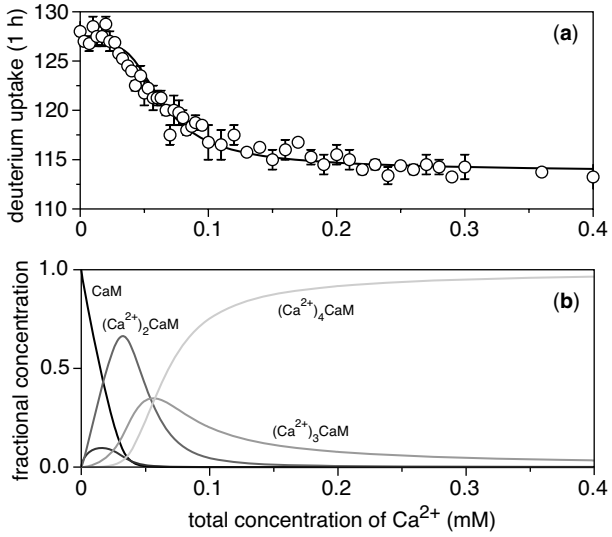


Figure 7.6. Use of PLIMSTEX to probe Ca^{2+} binding to porcine calmodulin. (*a*): The Ca^{2+} titration of 15 μM porcine calmodulin (CaM) in 50 mM *N*-(2-hydroxyethyl) piperozine-*N*-ethane sulfonic acid (HEPES) (pH 7.4, 21.5°C, 90% D_2O). (*b*) Fractional concentrations of $(Ca^{2+})_i$CaM species as a function of total Ca^{2+} concentration in solution. The PLIMSTEX titration data (*a*) are nearly a mirror image of the of the $(Ca^{2+})_4$CaM fractional concentration (*b*), suggesting that $(Ca^{2+})_4$CaM formation plays a major role in determining the titration curve shape. None of the nonspecific binding of more than four Ca^{2+} ions is registered by PLIMSTEX, indicating that if further Ca^{2+}–CaM binding occurs, it does not cause any significant conformational changes in proteins. [Adapted with permission from (50). Copyright © 2003 American Chemical Society.]

Exchange). Another recent modification of this method, known as dilution strategy PLIMSTEX (dPLIMSTEX) can be used to quantitate binding affinity in large systems, such as antibody–antigen complexes (53).

7.3.3. Binding Revealed by Changes in Ligand Mobility

Another creative approach to characterizing protein–ligand interaction in solution with MS was introduced several years ago by Clark and Konermann (55). The technique is applicable in situations where the ligand's physical dimensions are much smaller compared to those of the ligand–protein complex, so that diffusion coefficients of the two species in solution will differ very significantly from one another. The diffusion coefficients are determined by measuring the spread of an initially sharp boundary between two solutions of different concentration in a laminar flow tube due to Taylor dispersion (56). In the absence of protein–ligand interaction, the measured ESI MS dispersion profiles show a gradual transition for the protein and a steep transition for the small ligand (Fig. 7.7). However, if a non-covalent complex is formed in solution, the dispersion profiles of the two species will be very similar, since diffusion of the small ligand is now limited by the slow Brownian motions of the protein.

Unlike the direct ESI MS based methods of probing noncovalent complexes, this technique does not rely on survival of noncovalent complexes upon their transition from solution to the gas phase. On the contrary, harsh conditions in the ESI interface are required to dissociate the complex ion in the gas phase, such that the dispersion profiles of its constituents can be monitored separately. Figure 7.8 illustrates application of this technique to study noncovalent heme–protein interactions in myoglobin under both native and denaturing conditions. This method has been sucessfully used to screen a small library of low molecular weight saccharides to identify those that bind to a protein target (57) and even to estimate their binding energy (58). Competition screening may also be useful in cases where mixtures of potential ligands include analytes that are not amenable to direct ESI MS detection or when studying noncovalent interactions involving two or more macromolecules.

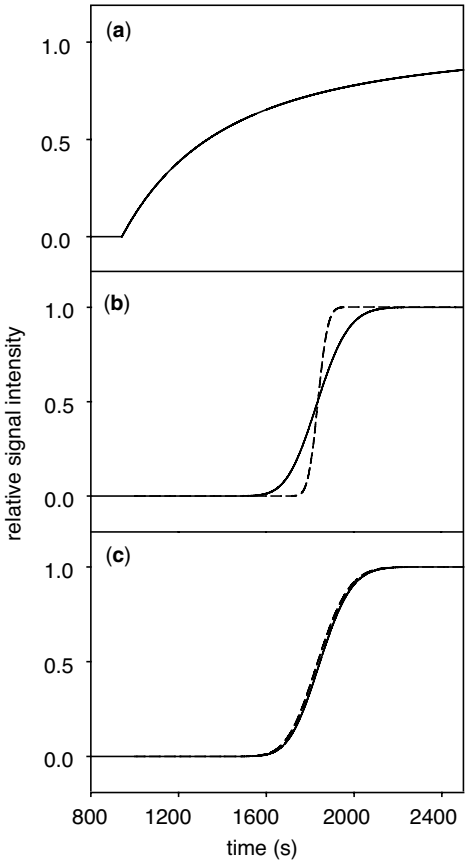


Figure 7.7. Simulated dispersion profiles representing signal intensity of selected analytes at the end of a laminar flow tube under different conditions. (*a*) A dispersion profile expected for a single analyte in the absence of diffusion. (*b*) Dispersion profiles expected for a protein ($D = 10^{-10}\,m^2/s$, solid curve), and a small ligand ($D = 10^{-9}\,m^2/s$, dashed curve) that do not interact in solution. (*c*) Dispersion profiles as in (*b*), but under the assumption of tight noncovalent binding between the two analytes (the profiles measured for the two species will be identical under these conditions). Flow parameters used in simulations: tube length, 3 m; tube radius, 129.1 μm; flow rate 5 μL/min (corresponding to $v_{max} = 3.18 \times 10^{-3}$ m/s). [Reproduced with permission from (55). Copyright © 2003 American Society for Mass Spectrometry.]

7.4. INDIRECT CHARACTERIZATION OF NONCOVALENT INTERACTIONS UNDER PARTIALLY DENATURING CONDITIONS

One of the methods of indirect assessment of protein–ligand binding considered in Section 7.3 is based on the assumption that ligand binding leads to greater stabilization of the native conformation of the protein. Such stabilization could only be detected with HDX, since both apo- and holo-forms of the protein are predominantly in the native conformation, with rare excursions to either partially or fully unfolded states. Ligand binding decreases the (already extremely low) frequency of unfolding events, thus making it possible to estimate binding energy as a difference between ΔG_{N^*} and ΔG_N (38). Under some circumstances, the stabilizing effect of ligand binding is much more pronounced, particularly if the apo-form of the protein fails to maintain the native conformation in the absence of ligand. In such cases, the stabilizing

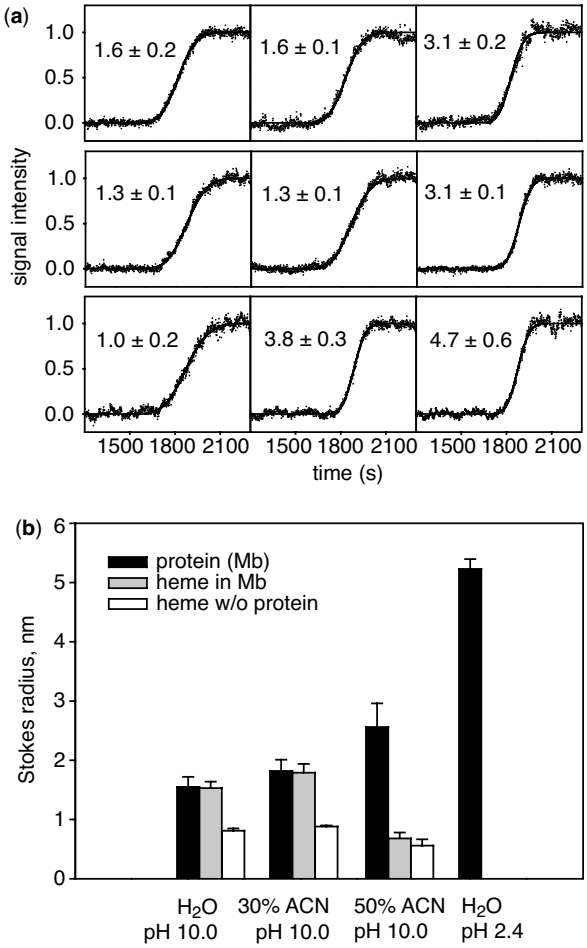


Figure 7.8. (*a*) Experimental dispersion profiles recorded for protein (first column) and heme (second column) in solutions of myoglobin (Mb) at pH 10. The third column shows dispersion profiles recorded for heme in protein-free solutions. The experiments were carried out in the presence of 0% acetonitrile (top row), 30% acetonitrile (middle row), and 50% acetonitrile (ACN) (bottom row). Solid lines are fits to the experimental data from which diffusion coefficients (shown each panel in units of 10^{-9} m^2/s). (*b*) Stokes radii of Mb and heme, calculated based on the experimentally determined diffusion coefficients (*a*). Problems with heme solubility precluded the determination of heme Stokes radius at pH 2.4. [Adapted with permission from (55). Copyright © 2003 American Society for Mass Spectrometry.]

effect of the ligand is so obvious that it does not require laborious HDX measurements, since the dramatic changes in large-scale protein dynamics induced by the ligand can be seen directly in the charge-state distributions of the protein ions in ESI mass spectra. Examples of such behavior (Fig. 7.9) abound particularly among metalloproteins, such as metallothioneins (4), endonucleases (59), or nuclear hormone receptor DNA binding domains (60).

7.4.1. Ligand-Induced Protein Stabilization Under Mildly Denaturing Conditions: Effect of Ligand Binding on Charge-State Distributions of Protein Ions

In most cases, however, both apo- and holo-forms of the protein are natively folded. Even in such circumstances, the stabilizing effect of the ligand can often be "felt" indirectly as it delays the onset of protein unfolding induced by mild chaotropes, when compared to the unfolding of the apo-form of the protein. We have seen such effects in iron-binding proteins from the transferrin family, where the presence of the ferric ion defers the acid-induced unfolding, as indicated by the protein ion charge-state distributions, by at least 1 pH unit (10,61). Importantly, metal ion dissociation from the holo-form of the protein in the absence of chelating agents is accompanied by a significant loss of tertiary structure, as indicated by a dramatic change in the protein ion charge-state distributions, strongly suggesting that there is an intimate link between metal binding and large-scale conformational dynamics.

The primary function of transferrin is sequestration and transport of iron in bodily fluids, but these proteins are also known to bind a variety of other metals. While for many of these metals binding to transferrin clearly manifests itself in ESI mass spectra as a mass increase, there is at least one metal ion (Bi^{3+}) whose binding to transferrin is known to occur, but cannot be verified in a straightforward ESI MS experiment (Fig. 7.10). This is a rather intriguing observation, since the Bi^{3+}–transferrin binding energy in solution is second only to that of Fe^{3+} (62). However, Bi^{3+}–transferrin binding can be easily detected indirectly by monitoring the evolution of charge state distributions of transferrin ions desorbed from a solution saturated with Bi^{3+} (Fig. 7.10). The onset of transferrin unfolding in metal-saturated solution occurs at the same pH 4.5, regardless of whether the metal is Fe^{3+}, In^{3+}, or Bi^{3+}. On the other hand, unfolding of the apo-protein occurs at significantly higher pH 5.5, providing a clear indication that Bi^{3+} does bind to transferrin in solution just like Fe^{3+} and In^{3+}. Apparently it dissociates from the protein very easily in the gas phase (unlike Fe^{3+} and In^{3+}), which makes direct detection impossible (62).

Although this indirect approach to detection of protein–ligand interaction has yet to be used widely, it has great potential. It is not as technically demanding as other indirect methods, and should be applicable to a wide spectrum of protein–ligand interactions. The major shortcoming is that such measurements only establish the fact of protein–ligand interactions taking place in solution without providing any information on binding stoichiometry or energetics. It is conceivable, though, that at least some semiquantitative data on binding energetics can be obtained in such measurements (e.g., ligand ranking according to their affinity), as stronger binding is expected to stabilize the protein against unfolding.

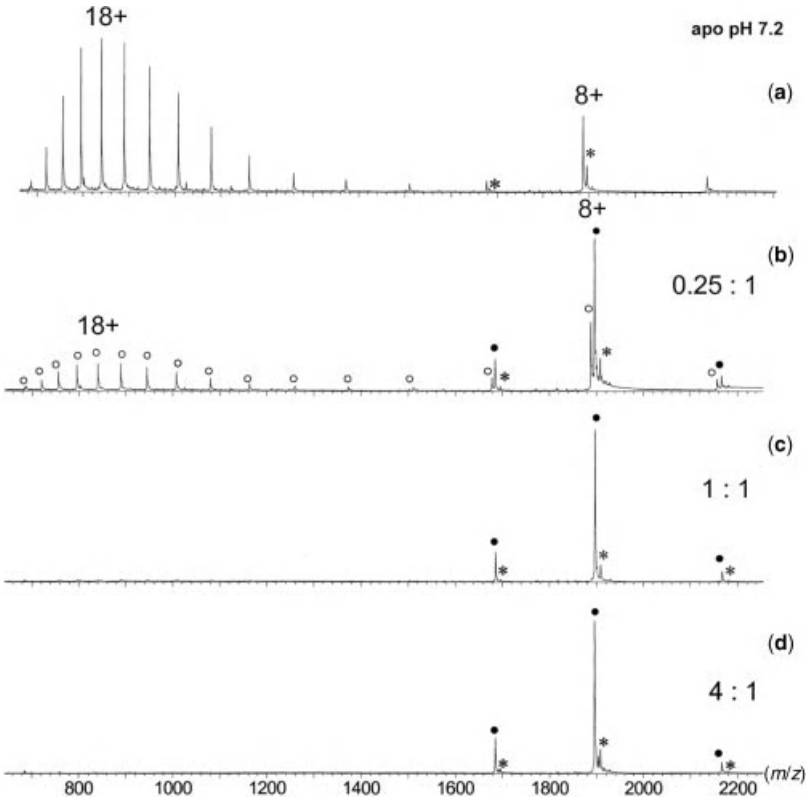


Figure 7.9. Stabilizing effect of Zn^{2+} on colicin E9 endonuclease (E9 DNase) conformation. Nano-ESI mass spectra of E9 DNase at pH 7.2 recorded in the absence (*a*) and in the presence (*b–d*) of Zn^{2+} in solution. Metal–protein molar ratios are 0.25 : 1 (*b*), 1 : 1 (*c*), and 4 : 1 (*d*). Ion peaks of Zn^{2+} bound and metal-free E9 DNase are labeled with filled and open circles, respectively. [Adapted with permission from (59). Copyright © 2002 The Protein Society.]

7.4.2. SUPREX: Utilizing HDX Under Denaturing Conditions to Discern Protein–Ligand Binding Parameters

This latter disadvantage (inability to provide quantitative affinity data) can be addressed in some cases using another indirect method of probing protein–ligand interactions, which relies on measurements carried out under denaturing conditions. This method, termed SUPREX (short for Stability of Unpurified Proteins from Rates of H/D EXchange), estimates protein stability in the presence of increasing amount of denaturant by measuring the extent of HDX at a certain time (63). Protein samples are subjected to HDX by dilution into a series of deuterated exchange buffers containing different concentrations of a denaturant (typically guanidinium chloride, GdmCl). After a specified exchange time, the deuterium content of each protein sample is determined with matrix-assisted laser desorption ionization (MALDI) MS. The change in mass (relative to the fully protonated sample) is plotted as a function of denaturant concentration to generate a SUPREX curve (Fig. 7.11).

The SUPREX curves can be used to extract thermodynamic parameters of protein unfolding, provided the equilibrium unfolding behavior conforms to a two-state process and that exchange occurs in the EX2 regime. As seen in Section 5.3, the exchange rate in a two-state system, where all labile amides are identical and no local fluctuations occur, can be expressed simply as:

$$k^{\mathrm{HDX}} = k_{\mathrm{int}}\, K \qquad (7\text{-}4\text{-}1)$$

where k_{int} is the "intrinsic" exchange rate (from the unprotected state) and K is the equilibrium constant for protein unfolding. Switching from K to the equilibrium constant of the reverse process K_{fold}, expression 7-4-1 can be rewritten as:

$$k^{\mathrm{HDX}} = k_{\mathrm{int}}/(1+K_{\mathrm{fold}}) \qquad (7\text{-}4\text{-}2)$$

As discussed in Section 5.3, k^{HDX} in Eq. 7-4-2 is defined in nuclear magnetic resonance (NMR) terms (*i.e.*, it is simply a rate of depletion of the number of protein

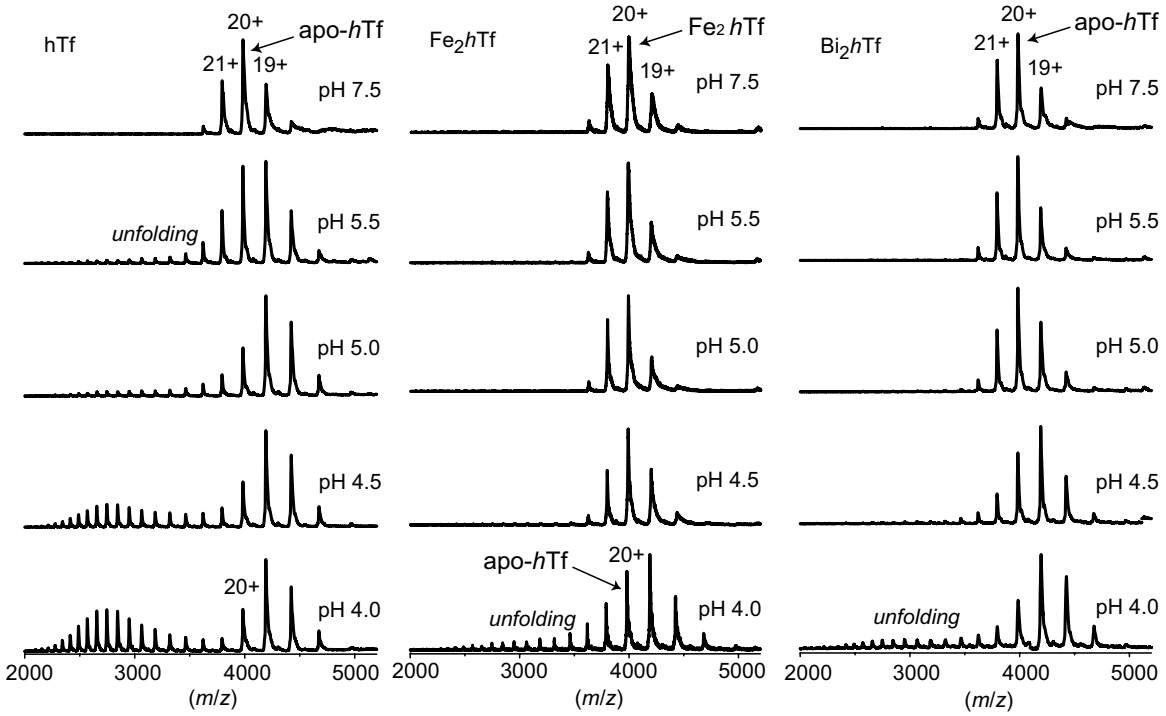


Figure 7.10. Acid-induced unfolding of *h*Tf monitored by ESI MS: apo-hTf (left column), Fe^{3+} saturated *h*Tf (center column), and Bi^{3+} saturated *h*Tf (right column). [Adapted with permission from (62). Copyright © 2004 American Society for Mass Spectrometry.]

molecules labeled with ^{1}H at a specific amide). On the other hand, HDX MS measurements regard this rate constant as a cumulative rate of exchange (an ensemble-averaged rate of loss of the entire ^{1}H content for all amides). If the basic SUPREX assumptions (as stated in the beginning of this paragraph) are correct, then both rate constants (measured by NMR and MS) should be numerically equal, although they actually have different meanings. The HDX progress in a two-state system is measured by MS simply as a protein mass gain ΔM:

$$\Delta M = \Delta M_\infty + (\Delta M_\mathrm{o} - \Delta M_\infty)\mathrm{e}^{-k^{\mathrm{HDX}} \cdot t} \qquad (7\text{-}4\text{-}3)$$

where ΔM_o is mass gain due to the exchange that does not require global unfolding, ΔM_∞ is the total mass gain after the exchange is complete, and t is the exchange time. Combining Eqs. 7-4-2 and 7-4-3, one can link the extent of exchange at any time t and the folding equilibrium constant K_{fold} as:

$$\Delta M = \Delta M_\infty + (\Delta M_\mathrm{o} - \Delta M_\infty)\mathrm{e}^{-(k_{\mathrm{int}}/1+K_{\mathrm{fold}}) \cdot t} \qquad (7\text{-}4\text{-}4)$$

The folding free energy of the folded state ($\Delta G = -RT \cdot \ln K_{\mathrm{fold}}$) in the presence of denaturant D is often estimated as:

$$\Delta G = \Delta G_\mathrm{o} + m\,[\mathrm{D}] \qquad (7\text{-}4\text{-}5)$$

where [D] is concentration of the denaturant, $m = \partial(\Delta G)/\partial[\mathrm{D}]$, and ΔG_o is the free energy of the folded state in the absence of denaturant. Therefore, if the parameters m and k_{int} are somehow known (or can be estimated), then the SUPREX curve (ΔM as a function of [D]) can be used to extract the value of ΔG_o.

Typically, a best fit to Eqs. 7-4-4 and 7-4-5 is found from a set of experimental data points (Fig. 7.11), from which a value of ΔG_o can be deduced by extrapolation to zero denaturant concentration. Although the SUPREX curves have the familiar look of conventional denaturation curves, they are actually rather different. The position of the midpoint in conventional denaturation curves $C_{1/2}$ is a function of ΔG_o and m, while in the SUPREX curves the position of the midpoint $C_{1/2}^{\mathrm{SUPREX}}$ also depends on the exchange time t and the intrinsic exchange rate k_{int} (63):

$$C_{1/2}^{\mathrm{SUPREX}} = C_{1/2} - \frac{RT}{m} \cdot \ln\left(\frac{k_{\mathrm{int}}\, t}{0.693} - 1\right) \qquad (7\text{-}4\text{-}6)$$

where $C_{1/2}$ can be used to calculate the free energy of folding simply as $\Delta G_o = -m\, C_{1/2}$. An extension of this method takes into account protein oligomerization in solution, which "locks" the protein in the exchange-incompetent state (65):

$$\Delta G_\mathrm{o} = -m \cdot C_{1/2} + RT \ln\left(\frac{n^\mathrm{n}}{2^{\mathrm{n}-1}}[\mathrm{P}]^{\mathrm{n}-1}\right) \qquad (7\text{-}4\text{-}7)$$

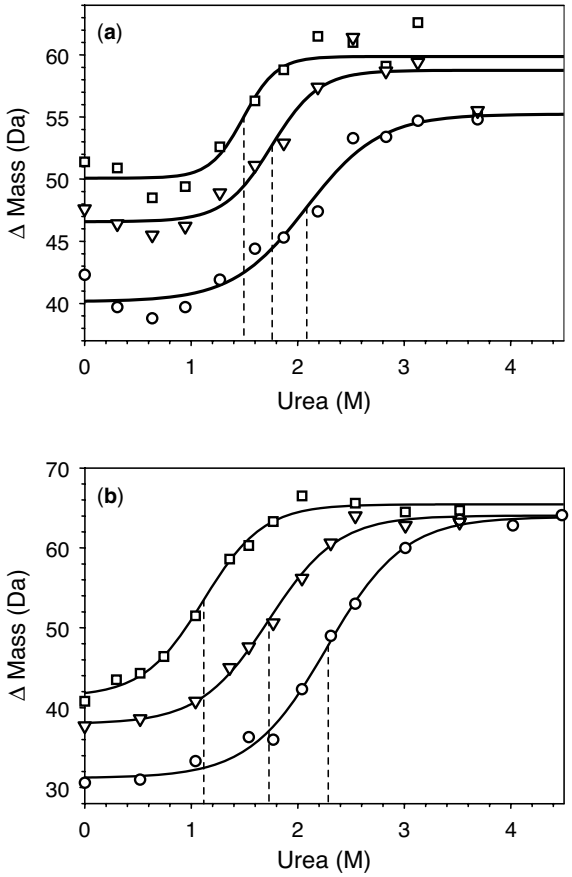


Figure 7.11. Representative SUPREX curves for S-Protein (10 μM) in the absence (*a*) and in the presence of 67 μM peptide 7 (*b*). (*a*) Exchange times of 10 (circles), 20 (triangles), and 30 min (squares) resulted in $C_{1/2}^{\text{SUPREX}}$ values of 2.08, 1.76, and 1.49 M urea concentration, respectively. (*b*) Exchange times of 25 (circles), 85 (triangles), and 312 min (squares) resulted in $C_{1/2}^{\text{SUPREX}}$ values of 2.28, 1.73, and 1.12 M urea concentration, respectively. The solid lines represent the best fits for each set of data. The vertical dotted lines mark the values of $C_{1/2}^{\text{SUPREX}}$. [Adapted with permission from (67). Copyright © 2003 American Chemical Society.]

where [P] is the protein concentration and n is the number of the protein subunits in the oligomer.

Since ligand binding usually stabilizes proteins, the ΔG_o values obtained in SUPREX measurements with and without ligands can be used to estimate protein–ligand binding energies (as $\Delta\Delta G_o$) and dissociation constants (66–68), as illustrated in Figure 7.11. One of the greatest advantages of the SUPREX methodology is its sensitivity and tolerance to a wide spectrum of salts, buffers, and other environmental conditions under which the HDX reactions and MALDI MS measurements can be carried out. It can also be automated and applied to screening small libraries (64). Unfortunately, the quality and reliability of global HDX measurements decreases precipitously when applied to large and/or heterogeneous proteins, which limits the scope of applications of SUPREX to relatively small systems. This limitation, however, may be overcome to a certain extent by incorporating local HDX measurements in the SUPREX routine (69).

Perhaps the major caveat of the quantitative analysis of SUPREX data (as is the case with most denaturation experiments) is that the protein must unfold in a cooperative, two-state process without any intermediate states present.* Another serious limitation of SUPREX is the requirement that HDX conform to the EX2 exchange regime, which is often impossible at medium or high denaturant concentrations or reasonably high pH, as discussed in Chapter 5.

7.5. UNDERSTANDING PROTEIN ACTION: MECHANISTIC INSIGHTS FROM THE ANALYSIS OF STRUCTURE AND DYNAMICS UNDER NATIVE CONDITIONS

The examples of using MS to characterize protein–ligand interactions considered so far have focused on identifying the players and extracting quantitative information regarding the strengths of such interactions in solution. Although such information is very important, in most cases it is not nearly sufficient to understand how proteins function, for example, to describe a mechanism of protein–ligand binding and identify structural and dynamic features of the protein–ligand system serving as a driving force of the interaction. The MS-based methods can often provide such mechanistic insights, allowing a diverse set of questions to be answered, ranging from catalytic activity of enzymes to long-distance signal transfer through allosteric interactions to induced protein folding. Specifically, exploitation of HDX under near-native conditions to locate and quantify functionally sensitive dynamic events often provides valuable mechanistic insights.

7.5.1. Dynamics at the Catalytic Site and Beyond: Understanding Enzyme Mechanism

The utility of HDX MS as a tool to probe functionally important structural and dynamic features of proteins was recognized early in its development. For example, mechanisms of enzymatic action at the atomic level are often deduced from the structures of the end-points of the reaction, that is, enzyme–substrate and enzyme–product complexes. Characterization of these complexes by HDX MS

*A multiphasic titration curve is observed (or else the transition is broadened) in such cases and any extrapolated stability measurement becomes impossible. However, the midpoint of the curve may still serve as a qualitative indicator of protein stability.

provides an important new dimension, as it often allows the distribution of protein flexibility throughout the entire protein to be determined. As discussed in Chapter 1, it is the conformational dynamics (not the structure alone) and the way it is distributed through the protein that often determines its biological activity. Therefore, probing enzyme dynamics in the presence and absence of substrates and/or products often provides invaluable information as to the molecular mechanism of its catalytic activity.

This approach is illustrated in work by Angeletti and co-workers (70), who used HDX MS to explore the effects of reduced nicotinamide adenine dinucleotide phosphate (NADPH) binding on the dynamics of *Corynebacterium glutamicum meso*-diaminopimelate dehydrogenase, a critical enzyme in the L-lysine biosynthetic pathway in bacteria. Binding of both NADPH and diaminopimelate to the enzyme was found to reduce the extent of amide exchange. Eight peptic fragments were identified whose deuterium exchange slowed considerably upon substrate binding, representing the regions known or thought to bind NADPH and diaminopimelate. One of the peptic fragments was located at the interdomain hinge region of the protein and was proposed to be exchangeable in the catalytically inactive open conformation, while remaining inaccessible to solvent in the catalytically active closed form of the enzyme formed after substrate binding and domain closure (70).

Structural changes remote from the active site were also found in another enzyme by Forest and co-workers (71), who used HDX MS to understand the mechanism of Mg^{2+} and NADPH induced activation of acetohydroxy acid isomeroreductase, the second enzyme of the parallel branched-chain amino acid pathway. The long-distance conformational changes were interpreted as structural effects of the ligand-binding process. Conformational changes not directly involved in the binding of substrates, observed by the same group in another system (creatine kinase isoenzyme) using HDX MS, were interpreted as a movement that allows the enzyme to properly align the substrates for optimal catalysis (72). Important mechanistic insights into the mode of enzymatic action can also be obtained by monitoring ligand-induced changes in HDX kinetics of enzyme analogues where mutations affect ligand-binding sites (73).

Detailed understanding of the mechanism of catalytic activity can be deduced from the analysis of conformation and dynamics of enzymes at various stages of the enzymatic reaction. This approach was exemplified in another work by the Angeletti group (74), who evaluated the dynamics of purine nucleoside phosphorylase at various stages of the catalytic cycle. Using HDX MS they were able to monitor the conformational differences within the enzyme induced by a substrate analogue, products, and a transition state analogue (enzyme inhibitor). Site-specific HDX measurements identified regions whose flexibility was decreased in all protein–ligand complexes. More importantly, several protein segments were identified whose dynamics were decreased only in complex with the inhibitor. The segments most highly influenced by the inhibitor surrounded the catalytic site, providing evidence for reduced-protein dynamic motion caused by the transition state analogue.

Figure 7.12 shows an example of using HDX MS to identify conformational changes important for the enzymatic action of kinases. In this work, carried out by Woods and co-workers (75), conformational dynamics were probed within the C-terminal Src kinase (Csk), a single kinase that down-regulates activity of all Src enzymes (a family of nonreceptor protein tyrosine kinases) by phosphorylating a single tyrosine residue at their C-termini. Hydrogen–deuterium exchange was carried out in the absence and presence of nucleotides ADP and the ATP analogue AMPPNP. A comparison of the ATP and ADP induced dynamic changes in the protein revealed unique structural changes induced by the γ-phosphate group of the nucleotide and provided a structural framework for understanding phosphorylation-driven conformational changes. Dynamics of various protein segments were monitored by measuring the deuterium content of the corresponding peptic fragments, allowing identification of segments that exchanged differentially with solvent (Fig. 7.12). Both ATP and ADP were found to protect similar regions of the protein, but the extent of such protection varied markedly in several crucial areas, such as the activation loop and helix G in the kinase domain, as well as several interdomain regions. The results of this study clearly demonstrated that delivery of the γ-phosphate group of ATP induced unique local and long-range conformational changes in Csk, which may have significant implications for regulatory motions in the catalytic pathway (75). A similar approach was later used to identify long-distance conformational changes transmitted through a cyclic adenosine 3′,5′-monophosphate (cAMP) dependent protein kinase upon its activation with cAMP (76,77).

Very often HDX MS allows detection of rather small, but functionally very important, conformational changes within large proteins. An example of such sensitive measurements was presented by Marshall and co-workers (78), who were able to detect and assign minute changes in the solvent exposure of troponin C induced by Ca^{2+} binding, which results in a very slight increase of the total exposed surface area. In many cases, however, the significant background of all exchangeable hydrogen atoms may interfere with the identification of those that are functionally relevant. This problem can be circumvented by using a combination of HDX MS and functional labeling, as discussed in Section 7.5.2.

In addition to identifying changes in conformation and dynamics of the enzyme induced by its interaction with substrates and inhibitors (79–91), or following activation by phosphorylation (92,93) or cofactor binding (94,95), HDX MS has been used to study structural features of enzymes,

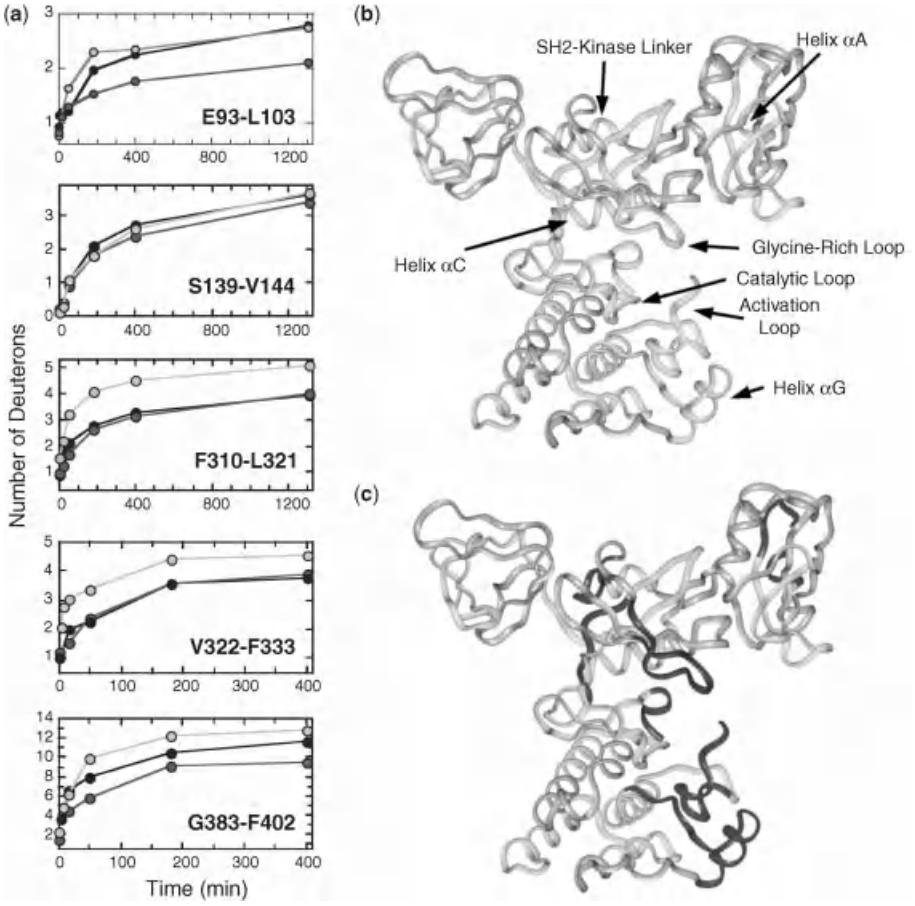


Figure 7.12. Probing phosphorylation-driven motions in the COOH-terminal Src kinase (Csk) with HDX MS. (*a*) Evolution of deuterium content of several peptic fragments of CsK in the absence (light gray) and presence of nucleotides: darker shade of gray, ATP analog AMPPNP and black, ADP. Coverage of Csk sequence with peptic fragments detected in HDX MS experiments (mapped to the X-ray structure for Csk) is shown in gray (*b*). The fragment peptides shown on panel (*a*) correspond to the following structural elements of Csk: E93-L103, A in SH2; S139-V144, SH2-kinase linker; F310-L321, catalytic loop; V322-F333, activation loop; and G383-G402, G. (*c*) HDX map of Csk showing effects of nucleotide binding: segments shown in darker shades of gray identify the regions whose protection is variably affected by ADP (as compared to ATP analog AMPPNP). Consult panels (*a*) and (*b*) for specific changes induced by each of the two ligands. [Adapted with permission from (75). Copyright © 2002 Elsevier Ltd.]

which often provide important clues vis-à-vis catalytic mechanisms (96–100). The HDX MS measurements can also provide important details on enzyme inactivation, when its higher order structure becomes compromised either locally or on a large scale due to inhibitor binding (101) or non-enzymatic post-translational modifications (102). At the same time, even though the lack of well-defined structure usually has a strong negative connotation in enzyme catalysis, HDX MS can also reveal instances where such a feature is vitally important for catalysis (e.g., by linking functional promiscuity with conformational heterogeneity in certain enzyme classes (103)).

In conclusion, note that any attempt to correlate structure, dynamics, and catalytic activity of an enzyme should always include careful consideration of time frames of all relevant processes. Most, if not all, enzymes act on the subsecond time scale, and so HDX measurements carried out on a longer time scale may not provide sufficient detail to characterize relevant catalytic processes in the active site. In recent work carried out in our laboratory with glucocerebrosidase (GCB, which is used in enzyme replacement therapy for a lysosomal disorder Gaucher's disease), stress oxidation was found to result in significant decrease of both biological activity of this protein as well as its conforma-

tional stability (102). Intriguingly, some of the conformational changes revealed by HDX MS were found to occur in the vicinity of the catalytic cavity of the protein. However, the time scale of these changes exceeded that of catalysis by several orders of magnitude, suggesting that catalytic function was not influenced directly by the oxidation-triggered changes in conformational dynamics. Instead, the relatively slow dynamic events leading to partial loss of GCB higher order structure are likely to facilitate aggregation and interfere with trafficking of this protein, thereby affecting its biological activity indirectly (102).

7.5.2. Allosteric Effects Probed by HDX MS

The ability of HDX MS to detect structural changes throughout the entire protein is particularly helpful in the studies of allosteric interactions (86,104–110). Although large allosteric changes are usually easy to spot in such measurements (provided the sequence coverage is adequate), a problem may arise when the magnitude of the conformational changes is small. This problem was elegantly solved by Englander et al., (111) who used tetrameric hemoglobin as a model system to study how protein molecules manage intramolecular signal transduction processes. Mammalian hemoglobins function by transducing a part of the binding energy of their initially bound O_2 ligands into structure-change energy. The energy is carried through the protein to distant heme sites in the form of structural changes, which are ultimately converted back into binding energy. Identification of individual allosterically important structural changes, their individual energetic contributions, as well as orchestration of such individual changes resulting in allosteric function were studied with HDX MS using the *functional labeling* technique (111,112). The idea of functional labeling as an extension of HDX measurements was initially discussed by Englander and Englander (113), namely, to focus only on those sites that change their exchange rates as a result of interactions. The experimental strategy is similar to the pulse-chase experiments commonly used in biology, and results in the isotope labeling only changing at sites affected by the folding or binding interaction. This eliminates a background of exchangeable amide hydrogen atoms that are located on the allosterically insensitive sites. In this manner, the HDX data can be much simplified, since complicating exchange processes due to local fluctuations are effectively removed from the measurements.

In the case of hemoglobin, amide exchange at the whole-molecule level shows significant differences between the oxy- (*R*-state) and deoxy- (*T*-state) forms. However, the number of allosterically sensitive amides and their exchange rates in either protein form cannot be determined because of the large background of unresolved sensitive and insensitive amides that exchange over a wide range of time scales (111). This difficulty is circumvented by using the functional labeling method, which focuses HDX labeling only on the sites that are affected by the $T \rightarrow R$ transition. The fast-exchanging *R*-state is partially labeled by exchange-in for a limited time period. Both allosterically sensitive and insensitive sites that exchange in this time period become labeled with D. The hemoglobin sample is then switched to the slow-exchanging *T*-state and transferred into protiated solvent, allowing the bound label (D) to exchange-out. After some exchange-out (chase) time, label on allosterically insensitive sites (same exchange rate in both forms) is largely lost, but it is retained on allosterically sensitive sites, which are more protected in the *T*-state. The exchange-in/exchange-out sequence produces a sample with the D label selectively placed on functionally sensitive sites. Site-specific HDX measurements then reveal the number of allosterically sensitive residues and their exchange rates in both forms of hemoglobin. This procedure allowed three sets of allosterically sensitive residues to be identified, which were then associated with known important structural changes occurring in hemoglobin upon the $T \rightarrow R$ transition (111).

7.6. GOING FULL CIRCLE WITH MS: NATIVE ESI MS REVEALS STRUCTURAL CHANGES PREDICTED BY HDX MS MEASUREMENTS

Although enzyme catalysis is one particular field where structural insights provided by MS-based methods (HDX MS in particular) shed light on the mechanism, there are many other situations where conformational changes revealed by this technique are critical for understanding the details of protein interactions with their partners. For example, mapping the interface between interacting proteins may yield, in addition to valuable structural information, evidence for dynamic events, such as folding or other conformational transitions that are critical for the binding process (114,115), while studies of conformational perturbations induced by binding of small signal molecules help understand protein activation processes (116). Understanding how protein conformation and dynamics may be affected by various factors (be it physical stimulus, noncovalent interaction with a physiological partner, or modification of the covalent structure) is also very helpful in many instances for pinpointing specific aspects of the function that will be influenced by such events. The example presented below demonstrates how this could be accomplished, but also goes beyond that by showing how orthogonal MS based methods can be used to verify these predictions.

The example given is the protein drug interferon-β1a (IFN), which is used for treatment of multiple sclerosis. As is the case with many biopharmaceuticals, IFN has a problematic propensity to misfold, which leads to activity loss, aggregation, and increased immunogenic response. This structure loss can be accelerated by a variety of factors, and some non-enzymatic PTMs have been long suspected to

act as triggers of misfolding. Specifically, alkylation of the sole free cysteine residue in this protein with *N*-ethylmaleimide (NEM) is known to lead to a significant reduction of biological activity, although the specific molecular mechanism leading to deactivation was not known. As seen in Chapter 5, both charge-state distribution analysis of IFN ions in ESI MS (Fig. 5.2) and HDX MS (Fig. 5.21) readily provide evidence of partial unfolding of NEM-modified IFN (117). Hydrogen–deuterium exchange MS is particularly helpful, as it identifies several segments in NEM IFN whose stability is greatly compromised by alkylation of Cys^{17} (117). One example is shown in Figure 5.21, where very slow uptake of deuterium is observed in a peptic fragment [L88–L102] derived from intact IFN, while protein alkylation results in a dramatic acceleration of the exchange kinetics of this segment. Although this protein segment is distant from the residue affected by alkylation (Cys^{17}) within the primary sequence, it is proximal to the Cys^{17} side chain within the three-dimensional structure of the protein (Fig. 5.2).

A detailed analysis of the backbone flexibility maps of intact and NEM modified IFN (118) provides important clues regarding the molecular mechanism of inactivation of the modified form of IFN. This protein exerts its action by binding to the ectodomains of two transmembrane interferon receptors, IFNAR1 and IFNAR2. Assembly of this ternary complex in the extracellular space is required in order to bring the cytoplasmic domains of IFNAR1 and IFNAR2 into close proximity. This latter step leads to activation of the JAK–STAT signaling pathway, which triggers a convoluted sequence of events that regulate the activity of interferon-stimulated response element (ISRE) in the nucleus (119,120). The receptors IFNAR1 and IFNAR2 cannot interact directly with each other without IFN mediation, and the assembly of the ternary complex is believed to proceed via IFN binding to the high-affinity receptor IFNAR2, followed by recruitment of the low-affinity receptor IFNAR1 (119). The latter event is accompanied by a conformational change in the ectodomain of IFNAR1, which propagates to its cytoplasmic domain, thereby initializing signal transduction within the cell (121).

The IFNAR1 and IFNAR2 binding interfaces revealed by mutagenesis (122,123) are localized in two distinct parts of the IFN molecule (Fig. 7.13*a*). Comparison of this interface map with the diagram deduced from HDX MS measurements that localizes changes of backbone flexibility as a result of alkylation (Fig. 7.13*b*) suggests that nearly the entire IFN–IFNAR1 interface is destabilized by alkylation, while only few changes are observed in the IFN–IFNAR2 interface. Importantly, the IFNAR1 binding interface of IFN maps on to the most stable region of the surface of intact IFN (118); the loss of conformational stability in that region of the protein is expected to have grave consequences vis-à-vis its ability to bind IFNAR1. Therefore, conformational

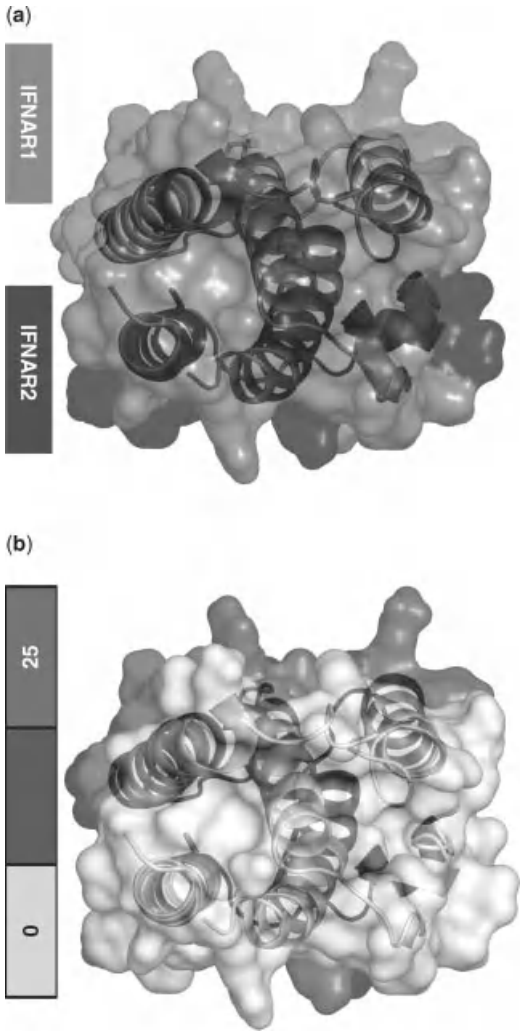


Figure 7.13. (*a*) Receptor-binding interfaces of IFN from earlier mutagenesis work (123). (*b*) Regions of IFN exhibiting a difference in backbone protection between the intact and NEM alkylated forms of IFN based on the results of HDX MS measurements (117). [Adapted with permission from (118). Copyright © 2010 American Society for Mass Spectrometry.] (See color version of the figure in Color Plate section)

consequences of IFN alkylation are likely to be highly detrimental to the IFN–IFNAR1 interaction, while the IFN–IFNAR2 interaction is not expected to be significantly affected. In this scenario, the loss of IFN biological activity as a result of its alkylation would be due to partial loss of its ability to bind one of the receptors (IFNAR1) and, consequently, form the ternary IFNAR1–IFN–IFNAR2 complex.

This hypothetical scenario can be easily verified (or refuted) by monitoring the efficiency of formation of protein–receptor complexes involving both intact and alkylated forms of IFN using the native ESI MS approach (118). Electrospray ionization MS of the IFN–IFNAR2 mixture acquired under near-native conditions provides clear

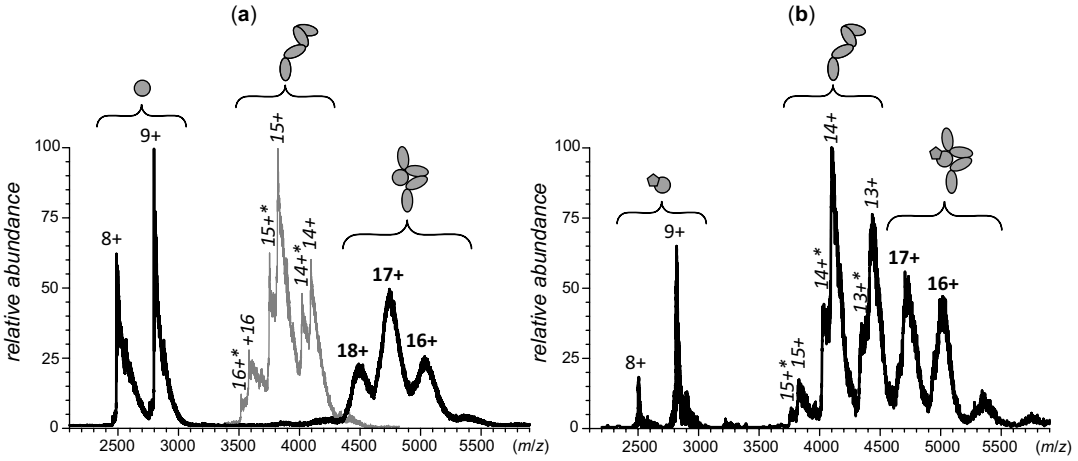


Figure 7.14. (*a*) Electrospray MS of IFN–IFNAR1 interaction in the presence of excess IFN. Note that free IFNAR1 is absent from the spectrum (a reference mass spectrum of free IFNAR1 is shown in gray). (*b*) Electrospray of NEM–IFN–IFNAR1 interaction in the presence of excess NEM–IFN. Note that both free NEM–IFN and IFNAR1 are present in the spectrum alongside the NEM–IFN–IFNAR1 complex. [Adapted with permission from (118). Copyright © 2010 American Society for Mass Spectrometry.]

evidence of a binary complex formation. The binding appears to be complete when the proteins are present in solution in the low micromolar concentration range, and does not appear to be disturbed by IFN alkylation (118). This finding is consistent with the predictions generated in the course of the analysis of HDX MS data presented in the previous paragraph. At the same time, alkylation of IFN results in significant weakening of its binding to the low-affinity receptor, as is evident from the ESI mass spectra of the IFN–IFNAR1 and NEM–IFN–IFNAR1 mixtures acquired under near-native conditions (Fig. 7.14). Not surprisingly, a very similar effect is exerted by IFN alkylation upon the stability of the ternary complex IFNAR1–IFN–IFNAR2 (118). It is very satisfying that native ESI MS confirms the conclusions drawn from the HDX MS experiments, regarding the consequences of PTM-triggered conformational changes for the ability of the protein to interact with its cognate receptors. Taken together, these two powerful techniques provide a means to determine the molecular basis of protein drug inactivation following deleterious PTMs or other stress-induced degradation events.

7.7. UNDERSTANDING PROTEIN ACTION: MECHANISTIC INSIGHTS FROM THE ANALYSIS OF STRUCTURE AND DYNAMICS UNDER NON-NATIVE (PARTIALLY DENATURING) CONDITIONS

The majority of the mechanistic studies focused on protein function discussed so far utilize HDX under native conditions. For most proteins, amide exchange under such conditions occurs in the EX2 regime, meaning that HDX MS experiments produce a picture of protein dynamics averaged across the entire protein population, just like HDX NMR. The existence of functionally important partially unstructured states (e.g., those participating directly or indirectly in the ligand-binding process) can be inferred from such experiments indirectly, that is, by grouping the amides exhibiting similar exchange rates. As seen in Section 5.3, switching the amide exchange mechanism to the EX1 regime allows, in many cases, distinct protein states to be identified, based on their respective residual levels of backbone protection. Unfortunately, switching exchange to the EX1 regime would also mean switching to denaturing conditions, raising doubts vis-à-vis relevance of any partially unstructured state(s) populated under such conditions to protein–ligand binding under native conditions. In this section, we will use CRABP I, a protein that has been already mentioned earlier in this chapter (Section 7.3.1), as a test case for addressing these concerns.

The CRABP I protein is a member of a large family of small, soluble intracellular proteins that bind hydrophobic ligands (e.g., fatty acids, lipids, and retinoids). This protein comprises 136 residues that form two five-stranded β-sheets. The first two strands are connected by a helix-turn-helix motif, and all of the others by reverse turns. The two β-sheets pack orthogonally to form a solvent-filled β-barrel. The ligand-binding pocket, which physiologically accommodates trans RA is located inside the barrel (Fig. 7.15). The ligand-binding–release mechanism by CRABP I has not been clearly established. A "portal" model has been postulated, which invokes the notion of a highly dynamic segment within the protein that serves as an opening, thus allowing the entry of ligand into the cavity (41). In CRABP I, this portal region constitutes the helix-turn-helix motif and two flanking

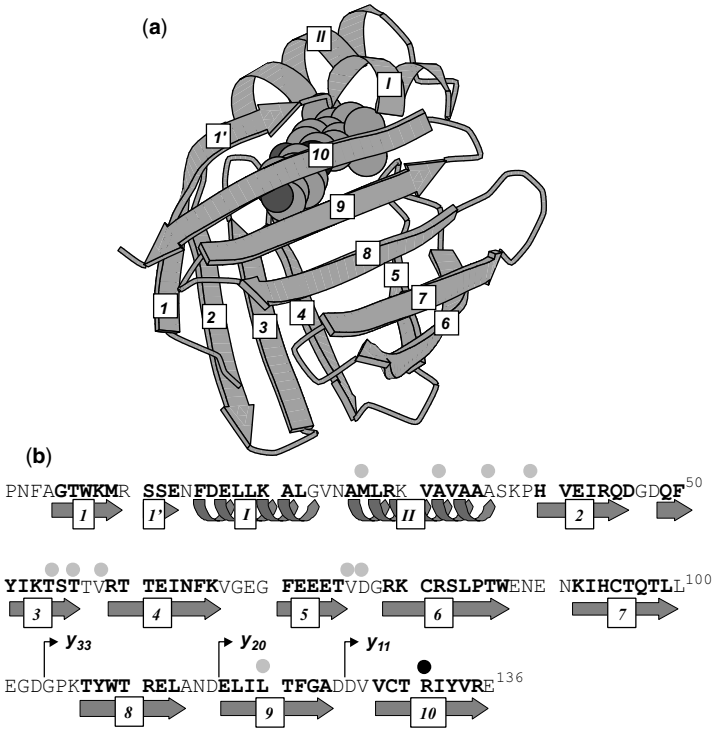


Figure 7.15. Ribbon diagram of CRABP I showing RA in the ligand-binding pocket (*a*) and the amino acid sequence of the protein with the X-ray derived ligand contact sites and secondary structural elements shown above and below the sequence, respectively (*b*). Gray circles indicate hydrophobic contacts; salt bridge-forming Arg residue is marked with a black circle.

β-hairpins. Reduced dynamics within this region in the holo form of the protein effectively inhibits dissociation of the complex in solution.

We saw earlier that the measurements of HDX kinetics of CRABP I carried out in aqueous solution at neutral pH (Fig. 7.5) indicate that the protein backbone dynamics is reduced dramatically in the presence of the cognate ligand (38), consistent with earlier NMR measurements. Although HDX MS kinetics measurements under native conditions can be used to obtain fairly accurate estimates of the RA–CRABP I binding energy, such measurements alone are not particularly revealing as far as the mechanism of protein–ligand binding is concerned. However, native HDX MS experiments do tell us that there are several intermediate states of the protein whose Boltzmann weight is affected by the presence of ligand (38). We previously postulated that at least some of these intermediate states, which are often viewed as "partial replicas of the native structure at lesser or greater degree of advancement"(124), are essential for ligand binding, as they provide easy access to the protein cavity. The transient nature of such intermediate states would prevent their direct observation under native conditions. Their presence, of course, can be detected in HDX MS and HDX NMR experiments. However, exchange follows the EX2 mechanism, making it impossible to differentiate among several possible activated states of the protein. To "visualize" the putative intermediate states of CRABP I that may play a role in ligand binding, we have to find a way to either increase their Boltzmann weight, which would affect the protein ion charge-state distributions, and/or to increase the potential energy barrier separating these activated states of the protein from the native conformation, which would affect the protein HDX by leading to longer residence times of the protein molecules within the potential wells corresponding to the intermediate states of the protein, and switching the HDX kinetics to the EX1 type. Both objectives can be achieved by inducing mildly denaturing conditions (e.g., mild acidification of the protein solution). Even though partial denaturation disrupts the protein–ligand interaction (ligand retention efficiency diminishes due to a decrease of the Boltzmann weight of the native state of the protein), some residual binding is observed until the pH is lowered to 3.0 (38).

The HDX kinetics clearly exhibit partial EX1 character within the pH range 3.0–3.5. The exchange pattern gives a clear indication that the presence of RA in the exchange buffer appears to stabilize the more structured states of the protein at a pH as low as 3.2. An example of such behavior is shown in Figure 7.16. In order to gain insight into the details of protein–ligand interaction, the isotope content of various fragment ions (produced by fragmenting CRABP I ions in the gas phase) was measured as a function of the exchange

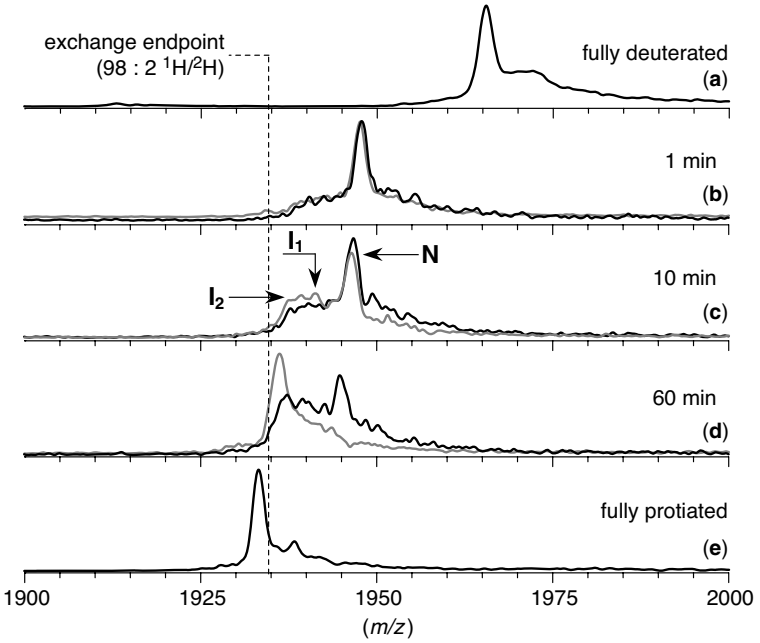


Figure 7.16. Global backbone dynamics of CRABP I probed with HDX–ESI MS under conditions favoring EX1 exchange mechanism (a 100 μM solution of a fully deuterated protein in D_2O was diluted 1:50, *v*:*v* in the exchange buffer solution, H_2O–$NH_4CH_3CO_2$, pH adjusted to 3.5 with CH_3CO_2H). Traces correspond to the fully deuterated (top) and unlabeled (bottom) protein; black and gray traces correspond to the exchange in the presence and in the absence of RA (5-fold excess) in the exchange buffer. All profiles correspond to a 8+ charge state of CRABP I ions.

time in solution. Protein ion fragmentation (CID) was carried out "in-source" (i.e., without mass selection of the precursor ion prior to dissociation) and was induced by increasing the capillary exit potential and ion residence time in the hexapole ion guide prior to injection to the ICR cell. Although full sequence coverage of CRABP I is not achieved under these conditions, the CID spectrum does contain an extended series of abundant structurally diagnostic fragment ions ranging in size from several amino acid residues to well over one-half of the protein. Figure 7.17 shows an example of time evolution of

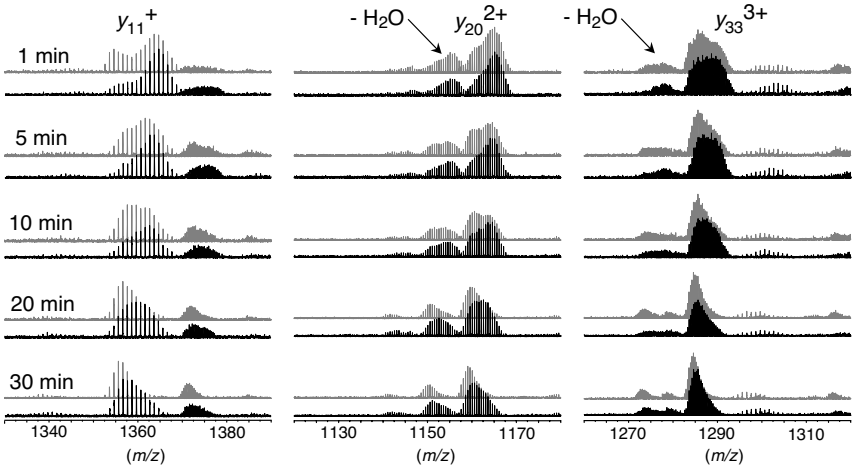


Figure 7.17. Effect of RA on the local backbone dynamics of CRABP I in aqueous acidic (pH 3.2) solution probed with HDX–ESI–CID MS. The three columns represent evolution of the isotopic distributions of three C-terminal fragments (y_{11}^+, y_{20}^{2+}, and y_{33}^{3+}) throughout 60 min of exposure of the deuterated protein to a protiated buffer (10 mM $NH_4CH_3CO_2$, pH adjusted to 3.2 with CH_3CO_2H). Less abundant clusters on the left-hand sides of the main peaks correspond to secondary fragments generated by a neutral loss of H_2O from the primary fragment ions.

the isotopic clusters corresponding to three C-terminal fragment ions of CRABP I with and without RA present in the exchange buffer. All three fragment ions (y_{11}^{+}, y_{20}^{2+}, and y_{33}^{3+}) contain Arg^{131}, a residue whose side chain forms a salt bridge with the carboxylate group of the ligand in the "native" complex, according to the available crystallographic data (125).

Consideration of the protein–ligand distances in the native complex suggests that Leu^{119} (contained within the y_{20}^{2+} and y_{33}^{3+} fragments) may form a hydrophobic contact with the ligand, while other amino acid residues within the segments under consideration do not appear to interact directly with RA. Isotopic clusters corresponding to all three fragments clearly exhibit bimodal character following short exposure to protiated solvent (top two traces in Fig. 7.17), indicating the presence of at least two protein states. This bimodal distribution is particularly clear in the case of the smallest fragment, y_{11}^{+}, with the two maxima separated by nine mass units. This separation gives a numerical value of the amide protection difference between the two conformations. Remarkably, this number is very close to a full number of backbone amides within the segment corresponding to the y_{11}^{+} fragment ion (i.e., 9 out of 10 amides remain deuterated following 1 min exposure to protiated solvent). The exchange pattern of this segment has familiar features of EX1 (cooperative) exchange overlaid with the exchange reactions due to uncooperative dynamic events, such as local structural fluctuations. Importantly, presence of the ligand clearly stabilizes the more structured conformation of the protein. Even after 30 min of protein exposure to protiated solvent under these denaturing conditions, the y_{11}^{+} fragment retains 2.5 ± 0.5 labile deuterium atoms when RA is present in the exchange buffer. All the while, the isotopic pattern of this fragment ion in the absence of the ligand is indistinguishable from that of a "fully exchanged" y_{11}^{+} fragment.

Similar analysis of the isotopic distribution of the y_{20}^{2+} and y_{33}^{3+} fragment ions as a function of exchange time in solution indicates that the fraction of protected backbone amides actually diminishes as the size of the protein segment under consideration increases. Thus, the "stable" protein form retains only 13 (out of possible 19) labile deuterium atoms within the [$Glu^{117} \rightarrow Glu^{136}$] segment following 1 min of exposure to the protiated solvent (as evidenced by the spacing between the two maxima in the isotopic clusters of the y_{20}^{2+} fragment ion). The increase in flexibility is even more dramatic within the larger segment, [$Gly^{104} \rightarrow Glu^{136}$] (only 17 out of 31 labile deuterium atoms are retained by the y_{33}^{3+} fragment ion; the isotopic clusters of the y_{33}^{3+} fragment ions appear to contain contributions from more than two species that cannot be clearly resolved). In our analysis, we take into account what appear to be the two extremes, the most and the least protected species. In both cases, presence of the ligand stabilizes the more protected form of the protein. Interestingly, the extent of measured protection at longer exchange times appears to be independent of the segment size. Thus, both y_{20}^{2+} and y_{33}^{3+} fragment ions retain only 3.0 ± 0.5 labile deuterium atoms in the presence of RA following 30 min exposure to the protiated buffer. This ligand-induced protection remains essentially the same (within experimental error) as for the much shorter segment corresponding to the y_{11}^{+} fragment ion.

These results clearly indicate that the acid-induced intermediate form of CRABP I interacts with the ligand at the "native" binding site located within β-strand 10 (Fig. 7.15). This conclusion is very important, as it proves that the equilibrium intermediate states of the protein populated under mildly denaturing conditions bind the ligand at the "native" site and, therefore, are likely to bear significant resemblance to the transient intermediates that facilitate ligand binding under native conditions. Similar analysis of the local exchange pattern deduced from the N-terminal (*b*-type) fragments representing strands 1, 1', and 2, and helices I and II also reveals uneven distribution of protein flexibility within the [$Pro^{1} \rightarrow Asp^{48}$] segment of the protein. Once again, the presence of RA in the exchange buffer dramatically affects local exchange kinetics, revealing intimate details of the ligand interaction with the activated (partially structured) states of the protein (126).

This chapter considered only a limited number of examples of various MS based approaches to study quantitative and mechanistic aspects of protein function. Because of space limitations, we could not possibly provide an exhaustive account of all work carried out in this field, so we focused on methodological aspects of such work. More examples of functional studies of various systems are presented in Chapter 9.

REFERENCES

1. Ganem, B., Li, Y. T., and Henion, J. D. (1991) Detection of noncovalent receptor ligand complexes by mass spectrometry, *J. Am. Chem. Soc. 113*, 6294–6296.
2. Ganem, B., Li, Y.-T., and Henion, J. D. (1991) Observation of noncovalent enzyme-substrate and enzyme-product complexes by ion-spray MS., *J. Am. Chem. Soc. 113*, 7818–7819.
3. Katta, V. and Chait, B. T. (1991) Observation of the heme globin complex in native myoglobin by electrospray-ionization mass-spectrometry, *J. Am. Chem. Soc. 113*, 8534–8535.
4. Yu, X., Wojciechowski, M., and Fenselau, C. (1993) Assessment of metals in reconstituted metallothioneins by electrospray mass spectrometry, *Anal. Chem. 65*, 1355–1359.
5. Hu, P., Ye, Q. Z., and Loo, J. A. (1994) Calcium stoichiometry determination for calcium binding proteins by electrospray ionization mass spectrometry, *Anal. Chem. 66*, 4190–4194.

6. Ganguly, A. K., Pramanik, B. N., Tsarbopoulos, A., Covey, T. R., Huang, E., and Fuhrman, S. A. (1992) Mass-spectrometric detection of the noncovalent GDP-bound conformational state of the human H-Ras protein, *J. Am. Chem. Soc. 114*, 6559–6560.
7. Cheng, X. H., Harms, A. C., Goudreau, P. N., Terwilliger, T. C., and Smith, R. D. (1996) Direct measurement of oligonucleotide binding stoichiometry of gene V protein by mass spectrometry, *Proc. Natl. Acad. Sci. USA 93*, 7022–7027.
8. Cheng, X., Morin, P. E., Harms, A. C., Bruce, J. E., Ben-David, Y., and Smith, R. D. (1996) Mass spectrometric characterization of sequence-specific complexes of DNA and transcription factor PU.1 DNA binding domain, *Anal. Biochem. 239*, 35–40.
9. Smith, R. D., Lightwahl, K. J., Winger, B. E., and Loo, J. A. (1992) Preservation of noncovalent associations in electrospray ionization mass spectrometry - Multiply charged polypeptide and protein dimers, *Org. Mass Spectrom. 27*, 811–821.
10. Gumerov, D. R. and Kaltashov, I. A. (2001) Dynamics of iron release from transferrin N-lobe studied by electrospray ionization mass spectrometry, *Anal. Chem. 73*, 2565–2570.
11. Halbrooks, P. J., Mason, A. B., Adams, T. E., Briggs, S. K., and Everse, S. J. (2004) The oxalate effect on release of iron from human serum transferrin explained, *J. Mol. Biol. 339*, 217–226.
12. Rondeau, P. and Bourdon, E. (2011) The glycation of albumin: Structural and functional impacts, *Biochimie 93*, 645–658.
13. Ghuman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A., Otagiri, M., and Curry, S. (2005) Structural basis of the drug-binding specificity of human serum albumin, *J. Mol. Biol. 353*, 38–52.
14. Kaltashov, I. A., Abzalimov, R. R., Eyles, S. J., and Frimpong, A. (2008) Studies of intact proteins and protein complexes: ESI MS approaches, In *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics* (Chance, M. R. Ed.), pp 215–240 Wiley–Blackwell, Boston.
15. Hofstadler, S. A. and Sannes-Lowery, K. A. (2006) Applications of ESI-MS in drug discovery: interrogation of non-covalent complexes, *Nat. Rev. Drug Discov. 5*, 585–595.
16. Hannah, V. V., Atmanene, C., Zeyer, D., Van Dorsselaer, A., and Sanglier-Cianferani, S. (2010) Native MS: an 'ESI' way to support structure- and fragment-based drug discovery, *Fut. Med. Chem. 2*, 35–50.
17. Huang, E. C., Pramanik, B. N., Tsarbopoulos, A., Reichert, P., Ganguly, A. K., Trotta, P. P., Nagabhushan, T. L., and Covey, T. R. (1993) Application of electrospray mass-spectrometry in probing protein-protein and protein ligand noncovalent interactions, *J. Am. Soc. Mass Spectrom. 4*, 624–630.
18. Light-Wahl, K. J., Schwartz, B. L., and Smith, R. D. (1994) Observation of the noncovalent quaternary associations of proteins by electrospray-ionization mass-spectrometry, *J. Am. Chem. Soc. 116*, 5271–5278.
19. Hunter, C. L., Mauk, A. G., and Douglas, D. J. (1997) Dissociation of heme from myoglobin and cytochrome b_5: comparison of behavior in solution and the gas phase, *Biochemistry 36*, 1018–1025.
20. Aplin, R. T., Robinson, C. V., Schofield, C. J., and Westwood, N. J. (1994) Does the observation of noncovalent complexes between biomolecules by electrospray-ionization mass-spectrometry necessarily reflect specific solution interactions, *J. Chem. Soc. Chem. Comm. 1994*, 2415–2417.
21. Ding, J. and Anderegg, R. J. (1995) Specific and nonspecific dimer formation in the electrospray ionization mass spectrometry of oligonucleotides, *J. Am. Soc. Mass Spectrom. 6*, 159–164.
22. Lim, H. K., Hsieh, Y. L., Ganem, B., and Henion, J. (1995) Recognition of cell-wall peptide ligands by vancomycin group antibiotics - studies using ion-spray mass-spectrometry, *J. Mass Spectrom. 30*, 708–714.
23. Greig, M. J., Gaus, H., Cummins, L. L., Sasmor, H., and Griffey, R. H. (1995) Measurement of macromolecular binding using electrospray mass-spectrometry - determination of dissociation-constants for oligonucleotide-serum albumin complexes, *J. Am. Chem. Soc. 117*, 10765–10766.
24. Loo, J. A., Hu, P. F., McConnell, P., Mueller, W. T., Sawyer, T. K., and Thanabal, V. (1997) A study of Src SH2 domain protein-phosphopeptide binding interactions by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 8*, 234–243.
25. Jørgensen, T. J. D., Roepstorff, P., and Heck, A. J. R. (1998) Direct determination of solution binding constants for non-covalent complexes between bacterial cell wall peptide analogues and vancomycin group antibiotics by electrospray ionization mass spectrometry, *Anal. Chem. 70*, 4427–4432.
26. Jørgensen, T. J. D., Staroske, T., Roepstorff, P., Williams, D. H., and Heck, A. J. R. (1999) Subtle differences in molecular recognition between modified glycopeptide antibiotics and bacterial receptor peptides identified by electrospray ionization mass spectrometry, *J. Chem. Soc. Perkin Trans. 2 1999*, 1859–1863.
27. Nesatyy, V. J. (2002) Mass spectrometry evaluation of the solution and gas-phase binding properties of noncovalent protein complexes, *Int. J. Mass Spectrom. 221*, 147–161.
28. Bruins, A. P. and Cook, K. D. (2007) Electrospray ionization: Principles and instrumentation, In *The Ecyclopedia of Mass Spectrometry* (Gross, M. L., and Caprioli, R. M., Eds.), pp 415–421, Elsevier, Amsterdam, The Netherlands.
29. Gabelica, V., Galic, N., Rosu, F., Houssier, C., and De Pauw, E. (2003) Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry, *J. Mass Spectrom. 38*, 491–501.
30. Gabelica, V., Rosu, F., and De Pauw, E. (2009) A simple method to determine electrospray response factors of non-covalent complexes, *Anal. Chem. 81*, 6708–6715.
31. Liu, J. and Konermann, L. (2011) Protein-protein binding affinities in solution determined by electrospray mass spectrometry, *J. Am. Soc. Mass Spectrom. 22*, 408–417.
32. Wang, W., Kitova, E. N., and Klassen, J. S. (2003) Influence of solution and gas phase processes on protein-carbohydrate

binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem. 75*, 4945–4955.

33. Cheng, X. H., Chen, R. D., Bruce, J. E., Schwartz, B. L., Anderson, G. A., Hofstadler, S. A., Gale, D. C., Smith, R. D., Gao, J. M., Sigal, G. B., Mammen, M., and Whitesides, G. M. (1995) Using electrospray-ionization FT ICR mass-spectrometry to study competitive-binding of inhibitors to carbonic-anhydrase, *J. Am. Chem. Soc. 117*, 8859–8860.
34. Gao, J. M., Cheng, X. H., Chen, R. D., Sigal, G. B., Bruce, J. E., Schwartz, B. L., Hofstadler, S. A., Anderson, G. A., Smith, R. D., and Whitesides, G. M. (1996) Screening derivatized peptide libraries for tight binding inhibitors to carbonic anhydrase II by electrospray ionization mass spectrometry, *J. Med. Chem. 39*, 1949–1955.
35. Krishnaswamy, S. R., Williams, E. R., and Kirsch, J. F. (2006) Free energies of protein–protein association determined by electrospray ionization mass spectrometry correlate accurately with values obtained by solution methods, *Protein Sci. 15*, 1465–1475.
36. Leverence, R., Mason, A. B., and Kaltashov, I. A. (2010) Noncanonical interactions between serum transferrin and transferrin receptor evaluated with electrospray ionization mass spectrometry, *Proc. Natl. Acad. Sci. USA 107*, 8123–8128.
37. Wortmann, A., Jecklin, M. C., Touboul, D., Badertscher, M., and Zenobi, R. (2008) Binding constant determination of high-affinity protein–ligand complexes by electrospray ionization mass spectrometry and ligand competition, *J. Mass Spectrom. 43*, 600–608.
38. Xiao, H., Kaltashov, I. A., and Eyles, S. J. (2003) Indirect assessment of small hydrophobic ligand binding to a model protein using a combination of ESI MS and HDX/ESI MS, *J. Am. Soc. Mass Spectrom. 14*, 506–515.
39. Tsai, C.-J., Kumar, S., Ma, B., and Nussinov, R. (1999) Folding funnels, binding funnels, and protein function., *Protein Sci. 8*, 1181–1190.
40. Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. (2002) Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective, *Curr. Opin. Struct. Biol. 12*, 197–203.
41. Krishnan, V. V., Sukumar, M., Gierasch, L. M., and Cosman, M. (2000) Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding, *Biochemistry 39*, 9119–9129.
42. Lu, J., Lin, C. L., Tang, C., Ponder, J. W., Kao, J. L., Cistola, D. P., and Li, E. (2000) Binding of retinol induces changes in rat cellular retinol-binding protein II conformation and backbone dynamics, *J. Mol. Biol. 300*, 619–632.
43. Franzoni, L., Lucke, C., Perez, C., Cavazzini, D., Rademacher, M., Ludwig, C., Spisni, A., Rossi, G. L., and Ruterjans, H. (2002) Structure and backbone dynamics of apo- and holo-cellular retinol- binding protein in solution, *J. Biol. Chem. 4*, 4.
44. Stone, M. J. (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding, *Acc. Chem. Res. 34*, 379–388.
45. Yu, L., Zhu, C. X., Tse-Dinh, Y. C., and Fesik, S. W. (1996) Backbone dynamics of the C-terminal domain of Escherichia coli topoisomerase I in the absence and presence of single-stranded DNA, *Biochemistry 35*, 9661–9666.
46. Zidek, L., Novotny, M. V., and Stone, M. J. (1999) Increased protein backbone conformational entropy upon hydrophobic ligand binding, *Nat. Struct. Biol. 6*, 1118–1121.
47. Fayos, R., Melacini, G., Newlon, M. G., Burns, L., Scott, J. D., and Jennings, P. A. (2003) Induction of flexibility through protein-protein interactions, *J. Biol. Chem. 278*, 18581–18587.
48. Zhu, M. M., Rempel, D. L., Du, Z., and Gross, M. L. (2003) Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX, *J. Am. Chem. Soc. 125*, 5252–5253.
49. Zhu, M., Rempel, D., and Gross, M. (2004) Modeling data from titration, amide H/D exchange, and mass spectrometry to obtain protein-ligand binding constants, *J. Am. Soc. Mass. Spectrom. 15*, 388–397.
50. Zhu, M. M., Rempel, D. L., Zhao, J., Giblin, D. E., and Gross, M. L. (2003) Probing Ca^{2+}-induced conformational changes in porcine calmodulin by H/D exchange and ESI-MS: effect of cations and ionic strength, *Biochemistry 42*, 15388–15397.
51. Huang, R. Y. C., Rempel, D. L., and Gross, M. L. (2011) H/D exchange and PLIMSTEX determine the affinities and order of binding of Ca^{2+} with troponin c, *Biochemistry 50*, 5426–5435.
52. Sperry, J. B., Shi, X., Rempel, D. L., Nishimura, Y., Akashi, S., and Gross, M. L. (2008) A mass spectrometric approach to the study of DNA-binding proteins: Interaction of human TRF2 with telomeric DNA, *Biochemistry 47*, 1797–1807.
53. Tu, T. T., Dragusanu, M., Petre, B. A., Rempel, D. L., Przybylski, M., and Gross, M. L. (2010) Protein–peptide affinity determination using an H/D exchange dilution strategy: Application to antigen-antibody interactions, *J. Am. Soc. Mass Spectrom. 21*, 1660–1667.
54. Chitta, R., Rempel, D., Grayson, M., Remsen, E., and Gross, M. (2006) Application of SIMSTEX to oligomerization of insulin analogs and mutants, *J. Am. Soc. Mass. Spectrom. 17*, 1526–1534.
55. Clark, S. M., and Konermann, L. (2003) Diffusion measurements by electrospray mass spectrometry for studying solution-phase noncovalent interactions, *J. Am. Soc. Mass Spectrom. 14*, 430–441.
56. Clark, S. M., Leaist, D. G., and Konermann, L. (2002) Taylor dispersion monitored by electrospray mass spectrometry: a novel approach for studying diffusion in solution, *Rapid Commun. Mass Spectrom. 16*, 1454–1462.
57. Clark, S. M., and Konermann, L. (2004) Screening for noncovalent ligand-receptor interactions by electrospray ion-

ization mass spectrometry-based diffusion measurements, *Anal. Chem. 76*, 1257–1263.

58. Clark, S. M. and Konermann, L. (2004) Determination of ligand-protein dissociation constants by electrospray mass spectrometry-based diffusion measurements, *Anal. Chem. 76*, 7077–7083.
59. van den Bremer, E. T., Jiskoot, W., James, R., Moore, G. R., Kleanthous, C., Heck, A. J., and Maier, C. S. (2002) Probing metal ion binding and conformational properties of the colicin E9 endonuclease by electrospray ionization time-of-flight mass spectrometry, *Protein Sci. 11*, 1738–1752.
60. Low, L. Y., Hernandez, H., Robinson, C. V., O'Brien, R., Grossmann, J. G., Ladbury, J. E., and Luisi, B. (2002) Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains, *J. Mol. Biol. 319*, 87–106.
61. Gumerov, D. R., Mason, A. B., and Kaltashov, I. A. (2003) Interlobe communication in human serum transferrin: metal binding and conformational dynamics investigated by electrospray ionization mass spectrometry, *Biochemistry 42*, 5421–5428.
62. Zhang, M., Gumerov, D. R., Kaltashov, I. A., and Mason, A. B. (2004) Indirect detection of protein-metal binding: Interaction of serum transferrin with In^{3+} and Bi^{3+}, *J. Am. Soc. Mass Spectrom. 15*, 1658–1664.
63. Ghaemmaghami, S., Fitzgerald, M. C., and Oas, T. G. (2000) A quantitative, high-throughput screen for protein stability, *Proc. Natl. Acad. Sci. USA 97*, 8296–8301.
64. Dearmond, P. D., West, G. M., Anbalagan, V., Campa, M. J., Patz, E. F., and Fitzgerald, M. C. (2010) Discovery of novel cyclophilin A ligands using an H/D exchange- and mass spectrometry-based strategy, *J. Biomol. Screen. 15*, 1051–1062.
65. Powell, K. D., Wales, T. E., and Fitzgerald, M. C. (2002) Thermodynamic stability measurements on multimeric proteins using a new H/D exchange- and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry-based method, *Protein Sc.: Pub. Protein Soc. 11*, 841–851.
66. Powell, K. D., Ghaemmaghami, S., Wang, M. Z., Ma, L. Y., Oas, T. G., and Fitzgerald, M. C. (2002) A general mass spectrometry-based assay for the quantitation of protein-ligand binding interactions in solution, *J. Am. Chem. Soc. 124*, 10256–10257.
67. Ma, L. and Fitzgerald, M. C. (2003) A new H/D exchange- and mass spectrometry-based method for thermodynamic analysis of protein-DNA interactions, *Chem. Biol. 10*, 1205–1213.
68. Powell, K. D. and Fitzgerald, M. C. (2003) Accuracy and precision of a new H/D exchange- and mass spectrometry-based technique for measuring the thermodynamic properties of protein-peptide complexes, *Biochemistry 42*, 4962–4970.
69. Hopper, E. D., Pittman, A. M. C., Tucker, C. L., Campa, M. J., Patz, E. F., and Fitzgerald, M. C. (2009) Hydrogen/deuterium exchange- and protease digestion-based screening assay for protein–ligand binding detection, *Anal. Chem. 81*, 6860–6867.
70. Wang, F., Scapin, G., Blanchard, J. S., and Angeletti, R. H. (1998) Substrate binding and conformational changes of Clostridium glutamicum diaminopimelate dehydrogenase revealed by hydrogen/deuterium exchange and electrospray mass spectrometry, *Protein Sci. 7*, 293–299.
71. Halgand, F., Dumas, R., Biou, V., Andrieu, J. P., Thomazeau, K., Gagnon, J., Douce, R., and Forest, E. (1999) Characterization of the conformational changes of acetohydroxy acid isomeroreductase induced by the binding of Mg^{2+} ions, NADPH, and a competitive inhibitor, *Biochemistry 38*, 6025–6034.
72. Mazon, H., Marcillat, O., Forest, E., and Vial, C. (2003) Changes in MM-CK conformational mobility upon formation of the ADP-Mg^{2+}-NO_3-creatine transition state analogue complex as detected by hydrogen/deuterium exchange, *Biochemistry 42*, 13596–13604.
73. Wang, F., Li, W., Emmett, M. R., Hendrickson, C. L., Marshall, A. G., Zhang, Y. L., Wu, L., and Zhang, Z. Y. (1998) Conformational and dynamic changes of Yersinia protein tyrosine phosphatase induced by ligand binding and active site mutation and revealed by H/D exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Biochemistry 37*, 15289–15299.
74. Wang, F., Miles, R. W., Kicska, G., Nieves, E., Schramm, V. L., and Angeletti, R. H. (2000) Immucillin-H binding to purine nucleoside phosphorylase reduces dynamic solvent exchange, *Protein Sci. 9*, 1660–1668.
75. Hamuro, Y., Wong, L., Shaffer, J., Kim, J. S., Stranz, D. D., Jennings, P. A., Woods, J., Virgil L., and Adams, J. A. (2002) Phosphorylation driven motions in the COOH-terminal Src kinase, Csk, revealed through enhanced hydrogen–deuterium exchange and mass spectrometry (DXMS), *J. Mol. Biol. 323*, 871–881.
76. Hamuro, Y., Zawadzki, K. M., Kim, J. S., Stranz, D. D., Taylor, S. S., and Woods, J., Virgil L. (2003) Dynamics of cAPK Type IIβ activation revealed by enhanced amide H/^{2}H exchange mass spectrometry (DXMS), *J. Mol. Biol. 327*, 1065–1076.
77. Zawadzki, K. M., Hamuro, Y., Kim, J. S., Garrod, S., Stranz, D. D., Taylor, S. S., and Woods, V. L., Jr. (2003) Dissecting interdomain communication within cAPK regulatory subunit type IIβ using enhanced amide hydrogen/deuterium exchange mass spectrometry (DXMS), *Protein Sci. 12*, 1980–1990.
78. Wang, F., Li, W., Emmett, M. R., Marshall, A. G., Corson, D., and Sykes, B. D. (1999) Fourier transform ion cyclotron resonance mass spectrometric detection of small Ca^{2+}-induced conformational changes in the regulatory domain of human cardiac troponin C, *J. Am. Soc. Mass Spectrom. 10*, 703–710.
79. Busenlehner, L. S. and Armstrong, R. N. (2005) Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry, *Arch. Biochem. Biophys. 433*, 34–46.
80. Codreanu, S. G., Thompson, L. C., Hachey, D. L., Dirr, H. W., and Armstrong, R. N. (2005) Influence of the dimer

interface on glutathione transferase structure and dynamics revealed by amide H/D exchange mass spectrometry, *Biochemistry 44*, 10605–10612.

81. Zhou, B. and Zhang, Z. Y. (2007) Application of hydrogen/deuterium exchange mass spectrometry to study protein tyrosine phosphatase dynamics, ligand binding, and substrate specificity, *Methods 42*, 227–233.
82. Liu, Y.-H. and Konermann, L. (2008) Conformational dynamics of free and catalytically active thermolysin are indistinguishable by hydrogen/deuterium exchange mass spectrometry, *Biochemistry 47*, 6342–6351.
83. Lei, Y., Pawelek, P. D., and Powlowski, J. (2008) A shared binding site for $NAD^{(+)}$ and coenzyme a in an acetaldehyde dehydrogenase involved in bacterial degradation of aromatic compounds, *Biochemistry 47*, 6870–6882.
84. Marques, M. R., Vaso, A., Ruggiero, J., Fossey, M. A., Oliveira, J. S., Basso, L. A., dos Santos, D. S., de Azevedo, W. F., and Palma, M. S. (2008) Dynamics of glyphosate-induced conformational changes of *Mycobacterium tuberculosis* 5-enolpyruvylshikimate-3-phosphate synthase (EC 2.5.1.19) determined by hydrogen-deuterium exchange and electrospray mass spectrometry, *Biochemistry 47*, 7509–7522.
85. Wang, S. Z., Sura, G. R., Dangott, L. J., and Fitzpatrick, P. F. (2009) Identification by hydrogen/deuterium exchange of structural changes in tyrosine hydroxylase associated with regulation, *Biochemistry 48*, 4972–4979.
86. Frantom, P. A., Zhang, H. M., Emmett, M. R., Marshall, A. G., and Blanchard, J. S. (2009) Mapping of the allosteric network in the regulation of alpha-Isopropylmalate synthase from *Mycobacterium tuberculosis* by the feedback inhibitor L-leucine: Solution-phase H/D exchange monitored by FT–ICR mass spectrometry, *Biochemistry 48*, 7457–7464.
87. Maegawa, G. H. B., Tropak, M. B., Buttner, J. D., Rigat, B. A., Fuller, M., Pandit, D., Tang, L. I., Kornhaber, G. J., Hamuro, Y., Clarke, J. T. R., and Mahuran, D. J. (2009) Identification and characterization of ambroxol as an enzyme enhancement agent for Gaucher disease, *J. Biol. Chem. 284*, 23502–23516.
88. Hsu, S., Kim, Y., Li, S., Durrant, E. S., Pace, R. M., Woods, V. L., and Gentry, M. S. (2009) Structural insights into glucan phosphatase dynamics using amide hydrogen-deuterium exchange mass spectrometry, *Biochemistry 48*, 9891–9902.
89. Edwards, A. A., Tipton, J. D., Brenowitz, M. D., Emmett, M. R., Marshall, A. G., Evans, G. B., Tyler, P. C., and Schramm, V. L. (2010) Conformational states of human purine nucleoside phosphorylase at rest, at work, and with transition state analogues, *Biochemistry 49*, 2058–2067.
90. Graf, C., Stankiewicz, M., Nikolay, R., and Mayer, M. P. (2010) Insights into the conformational dynamics of the E3 ubiquitin ligase CHIP in complex with chaperones and E2 enzymes, *Biochemistry 49*, 2121–2129.
91. Li, J., Dangott, L. J., and Fitzpatrick, P. F. (2010) Regulation of phenylalanine hydroxylase: Conformational changes upon phenylalanine binding detected by hydrogen/deuterium exchange and mass spectrometry, *Biochemistry 49*, 3327–3335.
92. Sours, K. M., Kwok, S. C., Rachidi, T., Lee, T., Ring, A., Hoofnagle, A. N., Resing, K. A., and Ahn, N. G. (2008) Hydrogen-exchange mass spectrometry reveals activation-induced changes in the conformational mobility of p38 alpha MAP kinase, *J. Mol. Biol. 379*, 1075–1093.
93. Hsu, Y. H., Johnson, D. A., and Traugh, J. A. (2008) Analysis of conformational changes during activation of protein kinase Pak2 by amide hydrogen/deuterium exchange, *J. Biol. Chem. 283*, 36397–36405.
94. Alverdi, V., Mazon, H., Versluis, C., Hemrika, W., Esposito, G., van den Heuvel, R., Scholten, A., and Heck, A. J. R. (2008) cGMP-binding prepares PKG for substrate binding by disclosing the C-terminal domain, *J. Mol. Biol. 375*, 1380–1393.
95. Hsu, Y. H., Burke, J. E., Stephens, D. L., Deems, R. A., Li, S., Asmus, K. M., Woods, V. L., and Dennis, E. A. (2008) Calcium binding rigidifies the C2 domain and the intradomain interaction of GIVA phospholipase A(2) as revealed by hydrogen/deuterium exchange mass spectrometry, *J. Biol. Chem. 283*, 9820–9827.
96. Wong, L., Lieser, S. A., Miyashita, O., Miller, M., Tasken, K., Onuchic, J. N., Adams, J. A., Woods, V. L., and Jennings, P. A. (2005) Coupled motions in the SH2 and kinase domains of Csk control Src phosphorylation, *J. Mol. Biol. 351*, 131–143.
97. Quint, P., Ayala, I., Busby, S. A., Chalmers, M. J., Griffin, P. R., Rocca, J., Nick, H. S., and Silverman, D. N. (2006) Structural mobility in human manganese superoxide dismutase, *Biochemistry 45*, 8209–8215.
98. Gu, Z. Y., Zitzewitz, J. A., and Matthews, C. R. (2007) Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: The roles of motif and sequence for the indole-3-glycerol phosphate synthase from Sulfolobus solfataricus, *J. Mol. Biol. 368*, 582–594.
99. Selevsek, N., Rival, S., Tholey, A., Heinzle, E., Heinz, U., Hemmingsen, L., and Adolph, H. W. (2009) Zinc ion-induced domain organization in metallo-beta-lactamases: A flexible "zinc arm" for rapid metal ion transfer?, *J. Biol. Chem. 284*, 16419–16431.
100. Seckler, J. M., Howard, K. J., Barkley, M. D., and Wintrode, P. L. (2009) Solution structural dynamics of HIV-1 reverse transcriptase heterodimer, *Biochemistry 48*, 7646–7655.
101. Coan, K. E. D., Maltby, D. A., Burlingame, A. L., and Shoichet, B. K. (2009) Promiscuous aggregate-based inhibitors promote enzyme unfolding, *J. Med. Chem. 52*, 2067–2075.
102. Bobst, C. E., Thomas, J. J., Salinas, P., Savickas, P., and Kaltashov, I. A. (2010) Impact of oxidation on protein therapeutics: Conformational dynamics of intact and oxidized acid-β-glucocerebrosidase at near-physiological pH, *Protein Sci. 19*, 2366–2378.
103. Hou, L., Honaker, M. T., Shireman, L. M., Balogh, L. M., Roberts, A. G., Ng, K. C., Nath, A., and Atkins, W. M. (2007) Functional promiscuity correlates with conformational heterogeneity in A-class glutathione S-transferases, *J. Biol. Chem. 282*, 23264–23274.

104. Rand, K. D., Andersen, M. D., Olsen, O. H., Jorgensen, T. J. D., Ostergaard, H., Jensen, O. N., Stennicke, H. R., and Persson, E. (2008) The origins of enhanced activity in factor VIIa analogs and the interplay between key allosteric sites revealed by hydrogen exchange mass spectrometry, *J. Biol. Chem. 283*, 13378–13387.

105. Laine, O., Streaker, E. D., Nabavi, M., Fenselau, C. C., and Beckett, D. (2008) Allosteric signaling in the biotin repressor occurs via local folding coupled to global dampening of protein dynamics, *J. Mol. Biol. 381*, 89–101.

106. Stokasimov, E. and Rubenstein, P. A. (2009) Actin isoform-specific conformational differences observed with hydrogen/deuterium exchange and mass spectrometry, *J. Biol. Chem. 284*, 25421–25430.

107. Li, S., Tsalkova, T., White, M. A., Mei, F. C., Liu, T., Wang, D., Woods, V. L., and Cheng, X. D. (2011) Mechanism of intracellular cAMP sensor Epac2 activation: cAMP-induced conformational changes identified by amide hydrogen/deuterium exchange mass spectrometry (DXMS), *J. Biol. Chem. 286*, 17889–17897.

108. Seckler, J. M., Barkley, M. D., and Wintrode, P. L. (2011) Allosteric suppression of HIV-1 reverse transcriptase structural dynamics upon inhibitor binding, *Biophys. J. 100*, 144–153.

109. Qi, Y. N., Chuah, M. L. C., Dong, X. M., Xie, K. L., Luo, Z., Tang, K., and Liang, Z. X. (2011) Binding of cyclic diguanylate in the non-catalytic EAL domain of FimX induces a long-range conformational change, *J. Biol. Chem. 286*, 2910–2917.

110. Iacob, R. E., Zhang, J. M., Gray, N. S., and Engen, J. R. (2011) Allosteric interactions between the myristate- and ATP-site of the Abl kinase, *PLoS ONE 6*, e15929.

111. Englander, J. J., Del Mar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y., and Woods, V. L., Jr., (2003) Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry, *Proc. Natl. Acad. Sci. USA 100*, 7057–7062.

112. Rogero, J. R., Englander, J. J., and Englander, S. W. (1986) Individual breathing reactions measured by functional labeling and hydrogen exchange methods, *Methods Enzymol. 131*, 508–517.

113. Englander, S. W., and Englander, J. J. (1994) Structure and energy change in hemoglobin by hydrogen exchange labeling., *Methods Enzymol. 232*, 26–42.

114. Sperry, J. B., Smith, C. L., Caparon, M. G., Ellenberger, T., and Gross, M. L. (2011) Mapping the protein–protein interface between a toxin and its cognate antitoxin from the bacterial pathogen streptococcus pyogenes, *Biochemistry 50*, 4038–4045.

115. Marcsisin, S. R. and Engen, J. R. (2011) Molecular insight into the conformational dynamics of the elongin BC complex and its interaction with HIV-1 Vif, *J. Mol. Biol. 402*, 892–904.

116. Chalmers, M. J., Busby, S. A., Pascal, B. D., West, G. M., and Griffin, P. R. (2011) Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions, *Exp. Rev. Proteom. 8*, 43–59.

117. Bobst, C. E., Abzalimov, R. R., Houde, D., Kloczewiak, M., Mhatre, R., Berkowitz, S. A., and Kaltashov, I. A. (2008) Detection and characterization of altered conformations of protein pharmaceuticals using complementary mass spectrometry-based approaches, *Anal. Chem. 80*, 7473–7481.

118. Kaltashov, I. A., Bobst, C. E., Abzalimov, R. R., Berkowitz, S. A., and Houde, D. (2010) Conformation and dynamics of biopharmaceuticals: transition of mass spectrometry-based tools from academe to industry, *J. Am. Soc. Mass Spectrom. 21*, 323–337.

119. Borden, E. C., Sen, G. C., Uze, G., Silverman, R. H., Ransohoff, R. M., Foster, G. R., and Stark, G. R. (2007) Interferons at age 50: past, current and future impact on biomedicine, *Nat. Rev. Drug Discov. 6*, 975–990.

120. Claudinon, J., Monier, M.-N., and Lamaze, C. (2007) Interfering with interferon receptor sorting and trafficking: Impact on signaling, *Biochimie 89*, 735–743.

121. Strunk, J. J., Gregor, I., Becker, Y., Li, Z., Gavutis, M., Jaks, E., Lamken, P., Walz, T., Enderlein, J., and Piehler, J. (2008) Ligand binding induces a conformational change in IFNAR1 that is propagated to its membrane-proximal domain, *J. Mol. Biol. 377*, 725–739.

122. Whitty, A. and Karpusas, M. (2003) The structure of human interferon-β-1a (Avonex^{TM}) and its relation to activity: a case study of the use of structural data in the arena of protein pharmaceuticals, In *Protein Structure: Determination, Analysis and Applications for Drug Discovery* (Chasman, D. L.,Ed.), pp 483–519, Marcel Dekker, New York-Basel.

123. Runkel, L., deDios, C., Karpusas, M., Betzenhauser, M., Muldowney, C., Zafari, M., Benjamin, C. D., Miller, S., Hochman, P. S., and Whitty, A. (2000) Systematic mutational mapping of sites on human interferon-β1a that are important for receptor binding and functional activity, *Biochemistry 39*, 2538–2551.

124. Englander, S. W., Mayne, L., and Rumbley, J. N. (2002) Submolecular cooperativity produces multi-state protein unfolding and refolding, *Biophys. Chem. 101–102* 57–65.

125. Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., and Jones, T. A. (1994) Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid, *Structure 2*, 1241–1258.

126. Kaltashov, I. A., Eyles, S. J., and Xiao, H. (2004) Combination of protein hydrogen exchange and tandem mass spectrometry as an emerging tool to probe protein structure, dynamics and function, In *Focus on Protein Research* (Robinson, J. W., Ed.), pp 191–218, Nova Science Publishers, Inc., Hauppauge, NY.

8

OTHER BIOPOLYMERS AND SYNTHETIC POLYMERS OF BIOLOGICAL INTEREST

So far, the majority of applications of mass spectrometry (MS) methods to study biomolecular structure and dynamics have focused on proteins. However, oligonucleotides and carbohydrates are also becoming targets of such studies. Unfortunately, many methods developed for proteins cannot be directly applied to study the dynamics of glycans and oligonucleotides (due to much higher rates of labile hydrogen exchange, etc.). In this chapter, we will focus primarily on several experimental methods developed specifically to probe the behavior of non-protein biopolymers. This discussion will be extended to consider uses of MS to study the behavior of polymers of nonbiological origin and their conjugates with biological molecules.

8.1. NUCLEIC ACIDS

8.1.1. Characterization of the Covalent Structure of Nucleic Acids

As the Human Genome Project got underway, many in the field suggested that MS would overtake traditional sequencing methodologies to become the method of choice for genomics studies (1–4). The ability of MS to distinguish species by mass at high resolution, the rapid acquisition time, and the excellent sensitivity of detection promised the possibility of reading off deoxyribonucleic acid (DNA) sequences directly in a matter of seconds, compared with the many hours required for electrophoretic separation. Short oligonucleotides were used as test molecules in the early developmental stages of both matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) (5,6). Although not as simple to achieve good data as with peptides or proteins, it appeared possible that MS would soon surpass traditional sequencing methods purely because of its speed. However, even approaching a quarter century since these initial experiments, this has not proved to be the case, for a variety of reasons.

The techniques for chemical sequencing of DNA are mature and highly sensitive (7,8). Since Sanger et al. (8) developed the technique for DNA sequencing using dideoxynucleotide chain termination, the process has become almost entirely automated. Originally the method involved mixing template DNA with a short ^{32}P radiolabeled primer sequence then adding DNA polymerase and nucleoside triphosphates (dNTP). Alternatively $\gamma(^{35}S)$-ATP (adenosine triphosphate) was employed to provide the radioactive tracer with the remaining three unlabeled NTPs to create a new strand complementary to the target sequence. After allowing time for chain extension, dideoxynucleotides (ddNTPs) were added to terminate extension of the DNA strand. This led to a mixture of DNA oligomers of varying lengths that contained radioactive label. The chains could then be separated by molecular weight using gel electrophoresis and the gel bands visualized by exposure to photographic film, where localized radioactive species would give rise to dark spots on the film from which the DNA sequence could be read. More recently this technique has been almost completely superseded in favor of automated sequencers using colored dye terminators that can be visualized fluorometrically with high sensitivity. Using ddNTPs with four different colored fluorophores, the DNA sequence of up to 1000 bases can be read directly from a single lane of a gel in a matter of several hours.

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

Modern sequencing methods forego the electrophoresis step completely and instead take advantage of reversible dye-based dNTP terminators. Primers attached to a glass slide are bound to the DNA template. At each extension cycle, the DNA is extended by a single nucleotide with the attachment of one of the four fluorescently terminated dNTPs, nonbinding species are removed, and an image is captured. Then the fluorescent moiety and the 3' blocking group are chemically removed and washed away before the next extension cycle. This method is currently one of the most commonly used for high-throughput DNA sequencing. Massively parallel methods that involve sequencing of vast numbers of short (25–50 nucleotide) read length DNA molecules and then reassembling these computationally has led to the recent availability of extensive genome coverage for an increasing number of organisms.

The above methods rely on duplication of the original DNA template and amplifying it using a DNA polymerase enzyme. While the fidelity of these enzymes is extremely high, there is the finite possibility of incorrect incorporation of nucleotides during synthesis that could lead to an incorrect (or at least ambiguous) sequence. On the other hand, mass spectrometric techniques can potentially be applied with high sensitivity to the original DNA template without the need for replication. Therefore this could remove the possibility of errors. One option is to analyze a mixture of products from the Sanger reaction by MS and read off the sequence in that manner, since acquisition of mass spectra is much more rapid than electrophoresis. An alternative and more attractive possibility is to measure intact DNA and obtain sequence information directly by tandem MS spectrometry (MS/MS) experiments (Fig. 8.1).

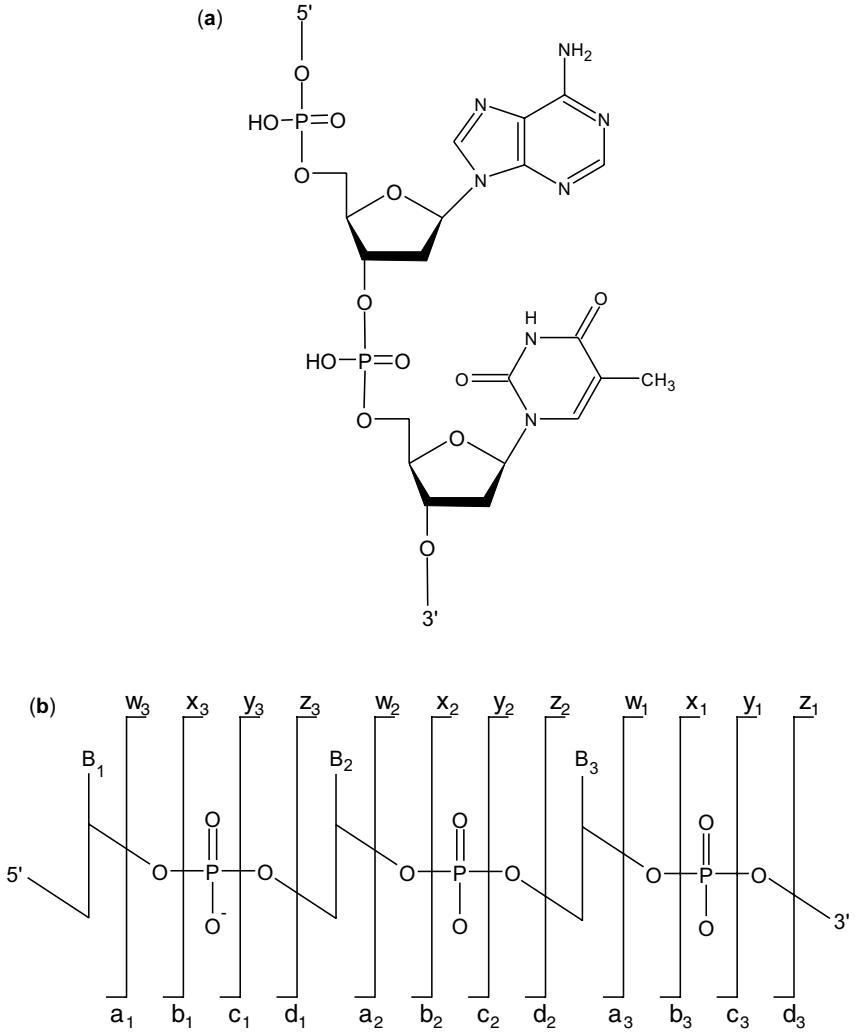


Figure 8.1. Structure of nucleotides showing the ribose or deoxyribose backbone of RNA and DNA with attached purine and pyrimidine nucleobases (*a*) and the MS/MS fragmentation nomenclature for nucleotides (*b*) proposed by McLuckey et al. (9).

The analysis of DNA by MS presents a number of complexities, not the least of which is sample preparation. The phosphodiester backbone of the molecule readily forms adducts with alkali and alkaline earth metal cations that remain intact in the gas phase, potentially leading to very broad peaks in the mass spectrum. This can be circumvented by buffer exchange into volatile ammonium salts that displace the metal cations and leave only the protonated (due to loss of neutral NH_3) species upon evaporation. Prior desalting by metal chelation or high-performance liquid chromatography (HPLC) has met with some success (10). An alternative approach employs gas-phase ion–ion reactions (11) to control the extent of adduction. This has been successfully applied using a dual spray setup of buffered oligonucleotide solution in one emitter and a chelating agent in the other, demonstrating that nonspecific metal adduction could be minimized by ion–ion reaction in the front-end quadrupole of a Fourier transform ion cyclotron resource (FT ICR) MS, giving a well-resolved spectrum (12).

Another problem is the ease of fragmentation along the phosphodiester backbone, that can lead to loss of signal from the intact DNA ion (13,14). This has the advantage of enabling a readout of limited sequence information from the MS1 spectra, but may prevent the possibility of a complete sequence read if there is insufficient ion intensity from intact species for subsequent MS/MS measurements. Other more disastrous fragmentation pathways (at least in terms of sequencing) involve removal of the base by 1,2-trans elimination, which render any fragmentation information meaningless. This may be circumvented by using a variety of chemical modifications of the DNA molecule prior to analysis in an attempt to reduce fragmentation effects (15–18).

Given these limitations, however, there have been significant advances in analysis of DNA by MS (19). Positive ion spectra have been obtained by MALDI of DNA molecules up to several hundred bases. It may be possible to minimize fragmentation caused by the laser irradiation and improve these mass limits by instead employing MALDI with lasers in the visible or infrared (IR) wavelength range using ice as the matrix (20–22). This produces a much more gentle ionization that may assist transportation of intact DNA into the gas phase. A more significant limitation of MALDI time-of-flight (TOF) MS for sequencing, however, is probably the mass discrimination effects and loss of resolution at higher m/z. Without a means to obtain a linear response through the high mass range and improved resolution, it is difficult to distinguish the identity of each base. The latter problem was investigated and determined to be primarily a problem of divergence of the ion beam for high mass ions, something that possibly may be solved by improvements in instrumental design (23). However, currently it is only practically possible to obtain unambiguous sequence information for relatively short oligonucleotides where high resolution is readily achievable.

On the other hand, negative ion ESI has shown promise for the analysis of larger DNA molecules since it alleviates the problem of adduct formation. The multiple charging of the ESI process brings ions into the mass spectral region where high resolution can be readily obtained. However, each species present produces many different charge states that can lead to significant complication of the spectra. Since a single species can give rise to a broad distribution of peaks, fragmentation either in-source or in MS/MS experiments would potentially give rise to a large number of peaks from various charge-state fragments that may be difficult to interpret. FT ICR shows significant promise in this regard due to its high resolving power that enables direct measurement of the charge state of fragment ions. In addition, there is strong evidence that, with carefully controlled source conditions, large DNA molecules can be effectively transferred intact into the gas phase for subsequent mass spectral analysis. Cheng et al. (24) were able to observe intact plasmid DNA (a circularized DNA molecule commonly used in molecular biology for cloning and protein expression in bacterial cells) using ESI FT ICR measurement, and to obtain mass measurement of a 1.95 MDa plasmid molecule within 0.2% mass accuracy, by employing charge reduction techniques using gas-phase reaction with acetic acid. Although this method may not be readily applicable to subsequent DNA sequencing, it demonstrates that MS is powerful for the analysis of polynucleotides, and further technological developments may yet bring MS to the forefront of DNA analysis.

Although MS cannot realistically compete with high-throughput or more traditional gel-based sequencing of DNA, it may excel at other applications when it comes to screening of differences in DNA primary structure. One of these is the detection of single nucleotide polymorphisms in genomes (25–28). Although the bulk of the DNA sequence of an organism is highly similar from one individual to another, there are cases where disease states may be caused by a single mutation at the genetic level. Higher organisms carry chromosomes in pairs within the cellular nuclei that duplicate the genetic information. While it is expected that the DNA sequence of the two paired chromosomes should be mostly identical, mutations occur at a rate of about one base per thousand. In many instances, due to parentage, a single nucleotide will differ between the two allelic sequences in an individual (known as heterozygous). If a single nucleotide is variable to a significant extent within a population, it is known as a single nucleotide polymorphism (SNP).* If this region of the DNA sequence is a coding region for a protein, then it will possibly lead to an amino acid mutation that may alter the structure and/or compromise function of the gene product, potentially lead-

* Generally this position is found to vary between just two of the four possible nucleotides.

ing to disease states (e.g., anemia, diabetes, or amyloidosis). Thus, the identification and analysis of SNPs has become a highly valued extension to the Human Genome Project, and it is here that MS excels due to the speed of analysis and ease of automation (29). Traditional DNA sequencing methods are at best ambiguous when it comes to the detection of SNPs, since the presence of two bands or fluorophores could be interpreted as experimental error rather than a true polymorphism. MS, on the other hand, can be employed to detect both species simultaneously, that differ in mass by that of the mutated nucleotide. The MALDI TOF MS has become one of the most widely used methods for determination of allele frequencies and scoring SNPs (30).

The same mass spectrometric techniques used for DNA sequencing can, in principle, be equally applied to determine the sequence of ribonucleic acid (RNA) molecules. Unfortunately, one complication is caused by the similarity in masses between the nucleosides uridine (306.2 Da) and cytidine (305.2 Da). Thus, a significantly higher mass resolving power is required to analyze longer RNA by standard MS methods. One way around this takes advantage of the differences in kinetics of phosphodiester hydrolysis by exonuclease enzymes, which sequentially remove bases from the ends of oligonucleotide polymers by hydrolysis of the phosphodiester bond for the two different bases, enabling ambiguities to be resolved by noting the different ion intensities (31). Bovine 5' → 3' phosphodiesterase hydrolyzes bonds adjacent to cytidine bases significantly more slowly than uridine, so this manifests itself in mass spectra as an increased ion intensity from fragments terminated with the former. Thus, sequence information could be obtained from enzymatic fragments of RNA generated by this enzyme. Of course, with high-resolution instrumentation, resolving this small mass difference is quite possible; thus, RNA sequencing by MS/MS measurements has since been achieved for small oligonucleotides (10,32). Additionally, infrared multiphoton dissociation (IRMPD) favors dissociation of oligonucleotides over peptides and proteins so this method has proven useful for sequencing in FT ICR MS (33).

Also of increasing interest in the field of RNA primary structure is the identification of post-translational modifications (PTMs). To date, 100 PTMs of RNA have been discovered, as well as various modifications to DNA (34). Using FT ICR MS in combination with fragmentation techniques (e.g., IRMPD), sequence ladders can be determined, as well as modifications based on the mass difference between peaks in the mass spectrum (35). Of particular interest is the recent development of synthetic RNA-cleaving deoxyribozymes that act in a manner analogous to proteolytic enzymes and can be engineered to have a selected residue and sequence specificity. These enable longer strands of RNA to be specifically cleaved into smaller segments, so that they are more amenable to analysis by MS and MS/MS (36).

8.1.2. DNA Higher Order Structure and Interactions with Physiological Partners and Therapeutics

Other than sequencing, there has been substantially less interest, at least until recently, in using MS for characterization of higher order structure of DNA molecules. Unlike proteins, or indeed RNA (see Section 8.1.3), DNA molecules were thought to adopt only relatively few favored conformations. The majority of naturally occurring DNA contains double-stranded dimers arranged in the classical Watson–Crick double helix. Additionally, most chromosomal DNA is further supercoiled with the aid of histone proteins to form compact structures that protect the genetic information maintained in them from being compromised by nucleases, regulate the on-demand transcription, and allow the genetic material to be confined within the small volume of the nucleus (Fig. 8.2). Structure and dynamics of this combination of DNA and proteins, called chromatin, obviously play a crucial role in many biological processes inherent to DNA (38). While the importance of the structure–plasticity balance of chromatin as a determinant of eukaryotic gene regulation is widely appreciated (38), many details remain elusive. In earlier chapters, we saw the power of HDX MS as a means of providing detailed information on quaternary structure of protein assemblies (Chapter 4), protein dynamics (Chapters 5 and 6), and their interactions with other biomolecules (Chapter 7). Unfortunately, this technology cannot be readily applied to study the higher order structure of nucleic acids, since the exchange proceeds on a much faster time scale (39). Nevertheless, HDX MS is capable of providing a great wealth of information on DNA packaging in chromatin,by focusing on the structure and dynamics of the protein counterparts of DNA, namely, histones (40,41).

Apart from the Watson–Crick double helical DNA structure, which is also known as the B-form DNA, a large number of other structures have been shown to exist, which either differ from the B conformation by arrangement of the two strands in the double helix (the so-called A and Z conformations), or by incorporating more than just two strands (e.g., triplexes and quadruplexes) (42). Several of these nonclassical DNA conformations came to prominence recently either due to their importance in designing novel therapeutic strategies (43) or for their potential use in nanotechnological applications (e.g., as scaffolds of building blocks in molecular devices) (44). One such conformation that has enjoyed significant attention in the past decade is the so-called G-quartet motif (Fig. 8.3) that evolves from 3' overhangs in the human telomeric sequence (5'-TTAGGG-3'), and forms a four-stranded structure (45).

We already saw in Section 4.1 that direct ESI MS measurements can allow noncovalent protein assemblies to be observed, provided that the solvent system is chosen in

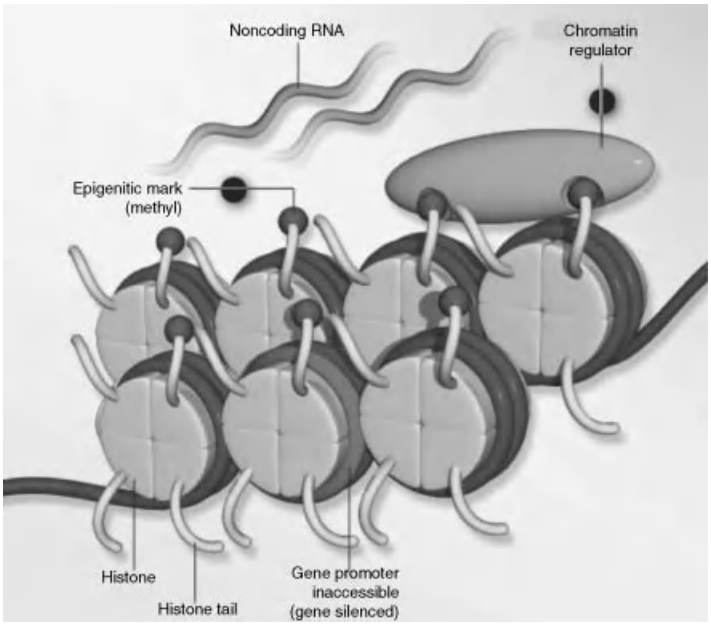


Figure 8.2. A schematic representation of The DNA packaging in chromatin. DNA double strands are wrapped around histones (shown as cylindrical structures). The strength of the binding is regulated by covalent modification of histone side chains (epigenetic marks) and chromatin regulators. [Adapted from (37). © 2011 The Scientist.] (See color version of the figure in Color Plate section)

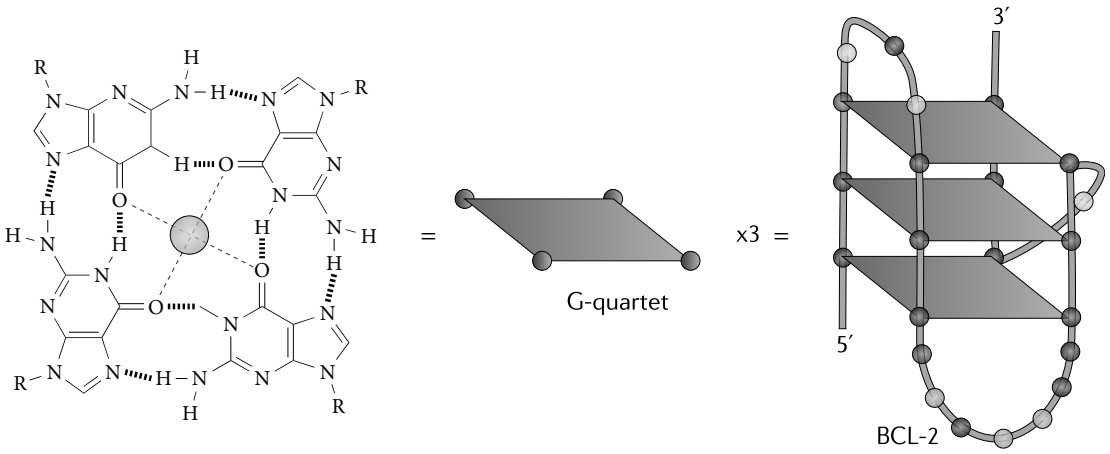


Figure 8.3. A schematic representation of a simple G-quadruplex structure built from the following components: a core of two or more π–π stacked G-quartets; essential alkali metal ions (Na^+ or K^+) that coordinate to O6 atoms of the guanine bases and are organized in a linear array running through the center of the core of G-quartets; and intervening variable-length sequences. The intervening variable-length sequences hold together the G-quartets and form loops that are arranged on the exterior of the core. The loops are the major elements that define structural variability in G-quadruplexes, and are analogous to side chains in amino acids. Even one short-length loop (<3 nucleotides) generally results in a G-quadruplex topology with at least two of the four backbones in a parallel orientation, and often with all four parallel to each other; this appears to be the dominant fold in most gene promoter G-quadruplexes examined to date. The variability in loop sequences results in highly variable (and sometimes flexible) cavities on the exterior of G-quadruplexes that can form part of ligand-binding sites. The example shown in the figure is for the dominant G-quadruplex structure that is found in the B cell lymphoma 2 promoter, in which the loop lengths are 3, 7, and 1 nucleotides. In this case, the small loop forms a parallel orientation and the two larger loops form antiparallel orientations. The stability of G-quadruplexes arises from the interplay between these various components. Thus, although other bases (e.g., thymine) can also form quartets, they are all significantly less stable than a G-quartet and, crucially, they are unable to effectively form the characteristic alkali metal ion channels. [Adapted from (43).]

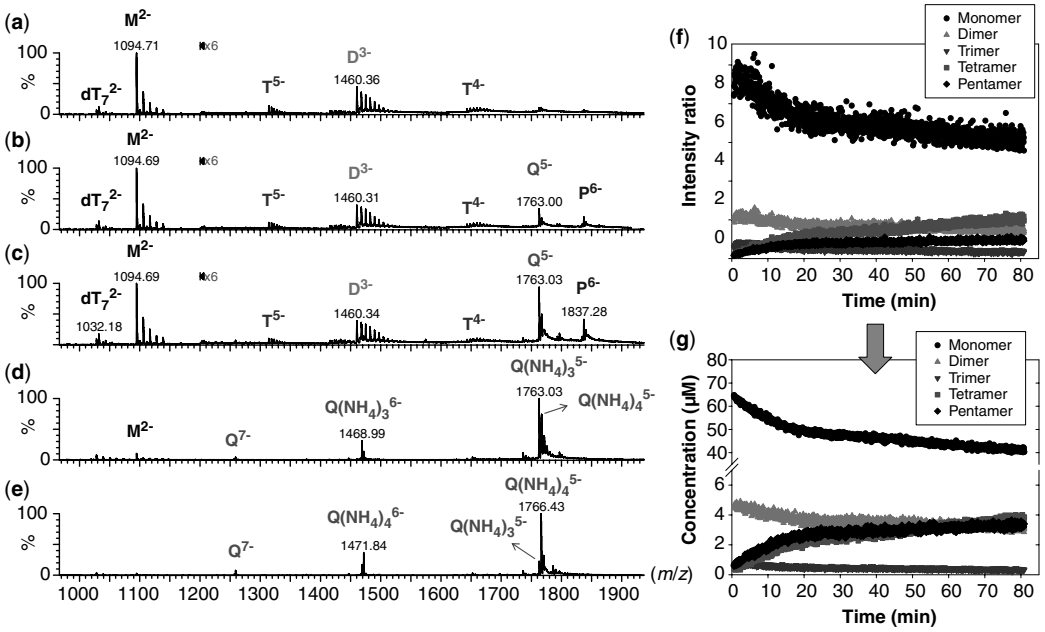


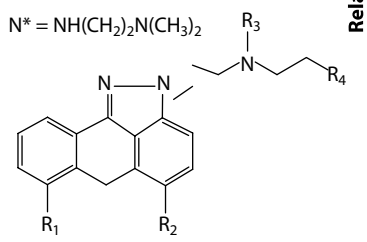
Figure 8.4. Kinetics of G-quadruplex formation by 80 mM dTG5T in 150 mM ammonium acetate and 10% methanol. Panels (*a*–*e*) show ESI mass spectra recorded 2 min (*a*), 10 min (*b*), 60 min (*c*), 25 days (*d*), and 164 days (*e*) after ammonium acetate addition. The parameter M stands for monomer, D is for dimer, T is for trimer, Q is for tetramer, and P is for pentamer. Note the magnification from 1200 to 2000 *m/z* in spectra (*a*–*c*). Panel (*f*) shows the intensity ratio measured between a characteristic peak of each stoichiometry (M^{2-} for monomer, D^{3-} for the dimer, T^{5-} for the trimer, Q^{5-} for the tetramer and P^{6-} for the pentamer), and the reference peak dT_7^{2-}, as a function of time elapsed after ammonium addition. Panel (*g*) shows time evolution of the concentration of each stoichiometry, obtained after correcting for the relative response factor of each characteristic peak. Note the break on the *y*-axis. [Reprinted from (49). Copyright © 2010 By permission of Oxford University Press.]

such a way as to be compatible with the ESI process, while at the same time causing minimal disruption to the macromolecular structure in solution. Earlier studies by Smith and co-workers (46) showed that this same approach could be used to obtain mass spectra of intact double-stranded DNA as well as tetramers of short oligonucleotides that assemble to form G-quadruplex-like structures (47). Formation of model G-quadruplexes was the focus of an extensive investigation by de Pauw and co-workers (48), who used direct ESI MS measurements to systematically study the influence of various factors on the stability of such structures. More recently, this group applied ESI MS to monitor the kinetics of formation of G-quadruplex structures from dTG_nT strands (Fig. 8.4), and used these data to propose a detailed mechanism of the tetramolecular complex formation (49).

Direct ESI MS measurements have also beensuccessful as a means of monitoring DNA interaction with small ligands, most notably DNA-targeting drugs. Following the initial observation by Smith and co-workers (50), numerous studies have been published where this technique was employed to evaluate not only the stoichiometry of such noncovalent complexes, but also their binding affinity [reviewed in (51,52)]. De Pauw and co-workers (see (51) for a detailed background of this method) used this technique to derive *absolute* values of drug–DNA binding constants based on relative intensity of ionic signals in ESI MS, an approach we briefly reviewed in Chapter 7. This method has become so successful that it was recently used as a screening tool to evaluate a library of novel triazole-based telomeric quadruplex-selective ligands that were developed to mimic an established family of tri-substituted acridine-based ligands, using crystal structure data as a starting point for computer-based design (53). A more conservative approach to evaluating DNA binding affinities of small-molecule drugs is exemplified by the work of Brodbelt [reviewed in (52)], who extracted *relative* binding affinities for a set of drug candidates from ESI MS data based on the fraction of ligand-bound DNA in the mass spectrum for each drug (Fig. 8.5).

So far, there have been relatively few studies targeting the structural aspects of DNA conformation beyond the stoichiometric measurements mentioned above. Bentzley and co-workers (54) attempted to evaluate the correlation between the extent of multiple charging of single-stranded and hairpin-forming short pieces of DNA in negative ion

Ligand	R1	R2	R3	R4
AP1	H	Cl	CH_3	OH
AP2	H	Cl	CH_3	Cl
AP3	H	Cl	H	OH
AP4	H	Cl	H	Cl
AP7	Cl	H	H	OH
AP9	Cl	H	CH_3	OH
AP10	H	N*	H	OH
AP11	H	N*	CH_3	OH
AP14	N*	H	CH_3	OH



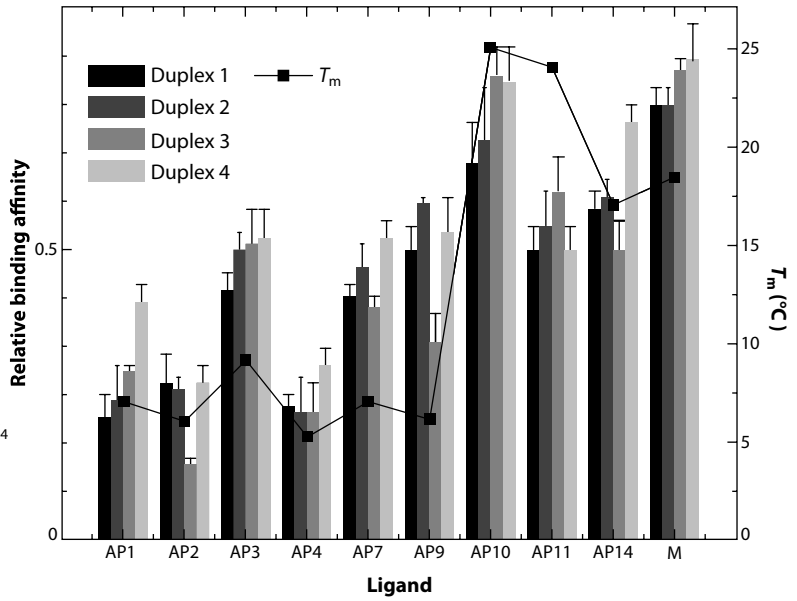


Figure 8.5. Structures of anthrapyrazoles (APs) and summary of relative binding affinities based on ESI MS measurements of fraction-bound DNA (bar graph) and the change in melting temperature (T_m) of duplex–AP complexes (right axis). [Abbreviation: M-mitoxantrone). [Reprinted with permission from (56). Copyright © 2007 John Wiley & Sons.]

ESI MS and reached a somewhat paradoxical conclusion that the ions representing linear structures exhibited lower charge density compared to the hairpin strands of the same composition. However, in our opinion this work lacked proper controls, and the charge-state distributions were extremely sensitive to a range of extrinsic factors. A more systematic evaluation of the utility of charge-state distributions of DNA polyanions in ESI MS was undertaken more recently by Touboul and Zenobi (55), who observed a strong correlation between the extent of multiple charging and the size of single-stranded deoxyoligonucleotides (Fig. 8.6). The conclusions of this work were consistent with the model of multiple charging of protein ions (57) that we discussed in Section 4.1.4. However, it remains to be seen if the extent

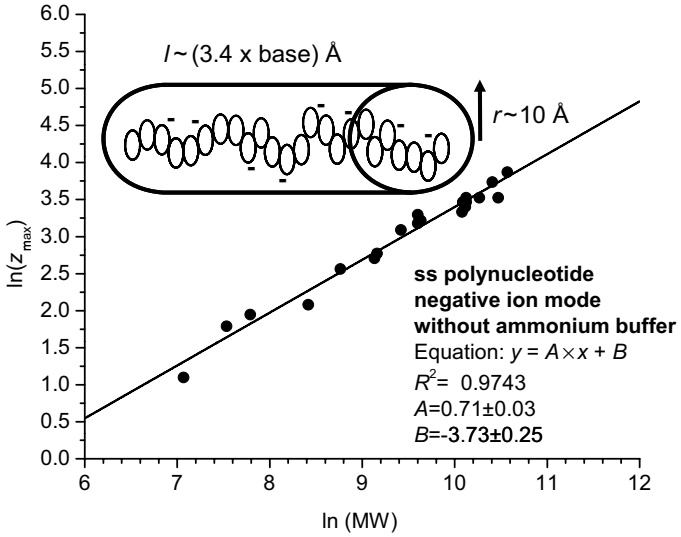


Figure 8.6. Maximum charge state z_{max} in negative ion ESI MS as a function of the molecular weight (MW) for single-stranded DNA (l = length, r = radius), plotted on a doubly logarithmic scale. From (56) [Reproduced by permission of the Royal Society of Chemistry.]

of multiple charging of DNA polyanions or polycations in ESI MS can be used to make a distinction among its various conformations.

Information on DNA higher order structure can also be provided by using selective chemical labeling and chemical cross-linking combined with MS analysis of the products, a technique similar to that discussed in Section 4.2 in conjunction with proteins. While a wide range of chemical probes of DNA structure are available (58), and the number continues to grow, MS has not been a prominent player in this field until recently. This situation is beginning to change, with the realization of the enormous potential of this technique as a tool to provide rapid and sensitive characterization of the reaction products of both cross-linking (59) and chemical labeling (60). Both of these techniques will be considered in detail in Section 8.1.3 in conjunction with RNA conformations.

We mentioned earlier in this section that hydrogen exchange of nucleic acids in solution is too rapid to provide meaningful data on the time scale accessible to traditional HDX MS measurements. One alternative to this approach utilizes gas-phase hydrogen exchange, since there is evidence that hydrogen bonding affords extra protection to labile protons, comparable to what would be expected in solution (61–63). This technique will be discussed in more detail in Section 8.1.3, where we consider various methods to probe the higher order structure of RNA.

Ion mobility (IM) is another gas-phase technique that has been evaluated in recent years as a means of providing structural information on DNA. However, as we already saw in Section 4.1, equating gas- and solution-phase structures of proteins may be a dangerous proposition; the same could be true for DNA. Bowers and co-workers (64) undertook a detailed study of gas-phase structures of DNA sequences capable of forming various conformations in solution. A careful analysis of the collisional cross-sections deduced from the IM–MS measurements was carried out in parallel with molecular dynamics calculations. The stoichiometry of noncovalent interstrand complexes detected in ESI MS matched those expected for various DNA conformations (hairpin, pseudo-knot, and cruciform). However, an agreement between the measured cross-sections and those calculated based on solution-phase structures was evident only for the lowest charge states of DNA polyanions (64), suggesting that electrostatic repulsion in the gas phase may lead to loss of the native conformation and favor elongated structures.

8.1.3. Higher Order Structure and Dynamics of RNA

The first part of the genetic code is generally considered to be DNA, from which messenger RNA (mRNA) is transcribed and subsequently translated into gene products (proteins). This is certainly true for higher organisms, but many viruses maintain all of their genetic code in the form of RNA. Upon injection of this RNA molecule into a host cell the viral ribonucleotide sequence is then treated like mRNA and translated into proteins that allow replication of the virus. Perhaps the most infamous of these RNA-only viruses are the class of retroviruses, including the human immunodeficiency virus (HIV) that carries not only RNA, but also copies of the enzyme reverse transcriptase. Upon host cell infection, the viral RNA is transcribed back into DNA and incorporated into the host genome, thus allowing the cellular machinery to be hijacked and produce large numbers of daughter viruses.

More recently, it was discovered that RNA is not merely a passive carrier of genetic information. The ribosome itself, the multiprotein complex machinery that produces proteins, also contains a number of RNA subunits that are functionally important. Additionally, several catalytic RNA sequences, known as ribozymes, have been identified. These are responsible for such tasks as splicing RNA sections together to modify the resulting gene product, and also for modifying transfer RNA (tRNA). Still other small RNA molecules may be responsible for regulating gene expression and a variety of other intracellular processes. Consequently, there has been much recent interest in identifying the structural and functional nature of these RNA molecules (65).

Unlike DNA, described above, which adopts relatively few different structures, a far greater range of structural features is available to RNA molecules. They are known to adopt a rich variety of secondary and tertiary interactions that make them extremely versatile, in spite of being composed of only four bases (cf. 20 different amino acids in proteins). This structural diversity is key to the wide range of functions that RNA is now known to perform. Ribozymes, the analogue of enzymes in the RNA world, must adopt conformations that enable them to specifically perform their catalytic function. Similarly, regulatory sequences must recognize their binding sites with high specificity. A whole class of small RNA molecules was identified, known as small interfering RNA (siRNA), a discovery that earned Fire and Mello (66) the Nobel Prize for Medicine in 2006. These small ($\sim$20 nucleotide) molecules can inhibit translation of other RNA molecules into protein, while others bind to chromosomal DNA and alter the genes that are exposed for translation. Consequently, there is a great deal of interest in determining the structural features of these RNA species and, in parallel with the protein-folding world, how these molecules attain their active conformations.

The biophysical tools for the study of RNA structure are still somewhat less mature than those for studies of proteins. For one thing, until now it was much more difficult to obtain pure RNA in significant quantity. Whereas proteins can be readily overexpressed in high yield in bacterial cells,

harnessing cells to produce large amounts of RNA is somewhat more complicated. *In vitro* transcription methods are now available that can produce the desired sequence of RNA in sufficient quantity for subsequent analysis. As with oligosaccharides, described in Section 8.2, RNA has been the subject of intense analysis by NMR. Unfortunately, there is relatively poor chemical diversity in oligonucleotides compared to polypeptides. The presence of aromatic nucleobases does, however, produce ring current shifts (NMR chemical shifts are influenced by the proximity of aromatic moieties that can either shield or deshield the nuclei and affect their resonant frequencies), so that there is reduced degeneracy of resonances in NMR spectra. Nevertheless, the protons are all either attached to the glycosidic backbone or to the purine or pyrimidine base, so there is generally much less dispersion than in a protein NMR spectrum, resonances instead being quite closely grouped in small regions of the spectrum (67–69). An additional problem is that nucleotides are much larger than amino acids, so relatively short RNA sequences can have molecular weights and consequently NMR correlation times reminiscent of large proteins, leading to unfavorable relaxation times and poor quality NMR spectra.

For similar reasons, high-resolution crystal structures have generally been reported only for relatively short RNA sequences, representing only domains of larger molecules. A major problem with RNA crystallography is the difficulty of producing isomorphous heavy atom derivatives for structure refinement (see Section 2.1). On the other hand, the crystal structure of the entire 70S ribosome was reported at 5.5 Å resolution (70), and of the separate subunits at still higher resolution (71,72). However, these are painstaking and time-consuming studies, thus other biophysical methods again emerge as key players to investigate structural features of RNA molecules. Once again, as with proteins, the static structure is informative, but still begs the question of how RNA folds into its biologically relevant structure, and what dynamic events are important for its function.

Hydrogen-deuterium exchange measurements have been employed to investigate the structure in RNA, using NMR (73,74), Raman spectroscopy (75), and $^{3}H^{-1}H$ exchange with radioactivity detection (76). Although the glycosidic hydrogens exchange rapidly, it is possible to measure protection of the base amino and imino protons that are involved in structure. This gives valuable information about base pairing as opposed to bases that are involved in single-stranded regions and/or bulges. Perhaps more useful in this respect is HDX in the gas phase, a method that shows promise for determining structural elements in oligonucleotides (62,77). Another technique for measuring structure due to hydrogen bonding is electron-detachment dissociation, the negative ion analogue of ECD (78). Removal of an electron from the RNA anionic species induces backbone fragmentation, but leaves hydrogen bonds intact (79), thus allowing regions of secondary structure to be inferred from "missing" fragments, and fragments ions that comprise intact double-stranded segments. This could-further be combined with other MS/MS techniques that disrupt hydrogen-bonding interactions to discriminate single- versus double-stranded sequences (36).

In Chapters 4 and 5, we described techniques using hydroxyl radical modification to probe protein structure. This method complements hydrogen exchange studies in proteins, but for the investigation of oligonucleotide structure and folding it is one of the few labeling methods that are truly effective. Exposure of folded RNA molecules to synchrotron radiation induces hydrolysis of the solvent-exposed backbone regions, thus cleaving regions that are on the outer surface of the molecule. Subsequent analysis by electrophoresis yields a ladder of fragments punctuated by missed cleavages that correspond to regions of structure protected against hydrolysis. This technique is known as hydroxyl radical footprinting (80,81). By measuring the degree of hydrolysis at equilibrium, structural changes (e.g., the folding of the *Tetrahymena* ribozyme as a function of Mg^{2+} concentration) could be determined (82). A further extension of this technique enabled measurement of the kinetics of RNA folding in response to addition of divalent cations, in a manner analogous to pulse-labeling experiments (83). Folding molecules are exposed to synchrotron radiation after a period of refolding, such that protection can be monitored as a function of folding time. In principle, MS detection may prove useful to determine the sites of cleavage. This methodology complements other low-resolution techniques to study RNA folding, such as small angle X-ray scattering.

Other chemical modifications can be employed to investigate RNA structure in a similar manner. A variety of reagents are available that act as solvent accessibility probes, since they cannot penetrate to modify, for instance, bases that are involved in base-pairing, stacking or other tertiary interactions. Using high-resolution FT ICR MS, Fabris and co-workers (84,85) demonstrated that these reagents [e.g., dimethyl sulfate, kethoxal, or 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimidemetho-*p*-toluenesulfonate] could be effectively used to probe RNA structure and RNA–protein interactions. The extent of chemical labeling of the structure was monitored by MS. Subsequent digestion with ribonuclease and analysis of the resulting fragments by high-resolution by MS enabled sites of modification to be localized (Fig. 8.7). In the case of multiple alkylation sites in a single fragment, these could be further analyzed by MS/MS. Other footprinting probes include *N*-methylisatoic anhydride, which acylates the 2'-OH group of nonbase-paired nucleotides, providing a measure not only of accessibility but also of structural flexibility in the RNA (86). These methods can be applied both to the study of the structure in RNA molecules, and also to interactions

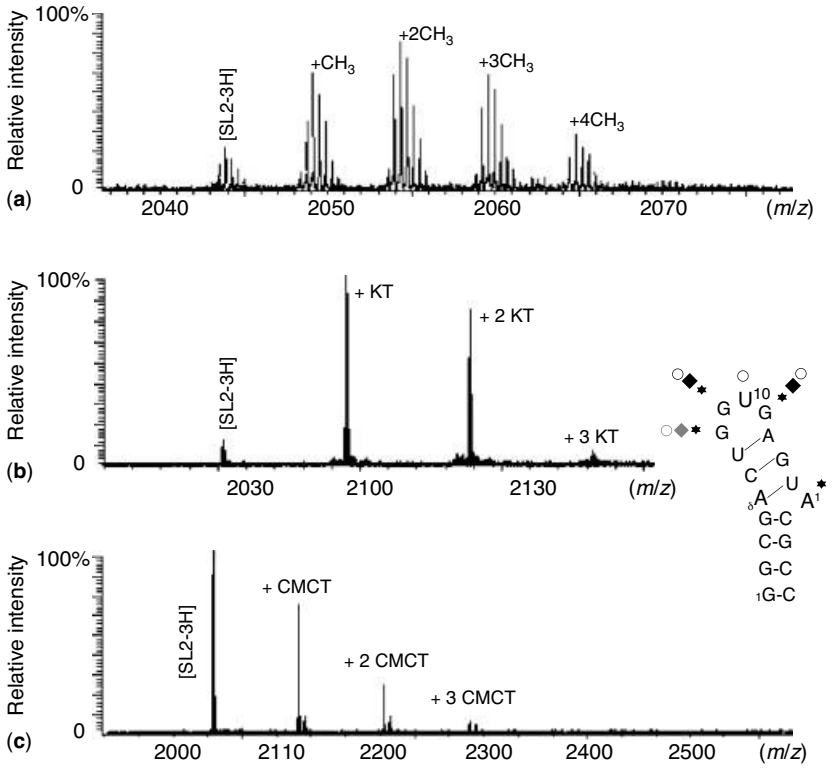


Figure 8.7. Structure in an RNA hairpin molecule (SL2 of the HIV-1 packaging signal) deduced from differential chemical modification. Only nucleobases that are not involved in intramolecular hydrogen bonding are available for modification by (*a*) methylation (CH_3) by dimethyl sulfate, (*b*) reaction with kethoxal (KT), or (*c*) 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimidemetho-*p*-toluenesulfonate (CMCT). [Reproduced from (84) Copyright © 2003 with permission from Elsevier Ltd.]

of oligonucleotides with RNA-binding proteins, since the binding event protects a previously exposed surface of the molecule against further chemical modification.

A technique, coined MS3D, for low-resolution identification of protein, folds from cross-linking experiments (87) (discussed in Section 4.2), has been elegantly adapted by the Fabris group (88) for probing the higher order structure of RNA, the workflow for which is shown in Fig. 8.8. Essentially, the structure of the polynucleotide under native conditions is probed by a series of chemical footprinting reagents. These solvent accessibility probes have varying specificity for different bases, and their reactivity is limited by the presence of base-pairing, stacking, or other tertiary interactions. Following labeling, the sites of modification are determined by a combination of bottom-up (digestion with ribonucleases) or top-down (gas-phase fragmentation) methods. Additional MS/MS techniques can be used to pinpoint the labeled site to the individual nucleotide. Combined with bifunctional cross-linking and photoactivatable cross-linkers, this methodology has been successfully employed to model 3D structures of RNA molecules. These methods can give distance constraints and information about the spatial arrangement at structural elements. Indeed it has proved possible to obtain a 3D structural model of the complete HIV-1 Ψ-RNA from contacts determined by cross-linking, combined with NMR structures of the individual stem-loop regions (90). Recent findings by the Fabris group (91–93) suggest that certain gas-phase fragmentation methods routinely used in FT ICR MS instruments may retain the nucleotide base-pairing even after activation, thus yielding fragments that have backbone cleavages, but intact base-pairing interactions. While this can be a limiting factor to obtaining complete top-down sequence coverage, the advantages of this method for identifying double-stranded regions of the RNA sequence may prove invaluable.

A natural extension of this methodology is to investigate binding of RNA to other molecules (e.g., proteins). Nucleotides that become buried at the interaction surface are rendered inaccessible to chemical-labeling reagents, thus the absence of modification indicates that the site is involved in binding. Similarly, complementary information can be obtained from cross-linking strategies to give RNA–protein distance constraints. This can also be combined with solution-phase HDX MS of the protein in the absence and presence of RNA binding to map out the interaction interface. Care must be taken, as with all of

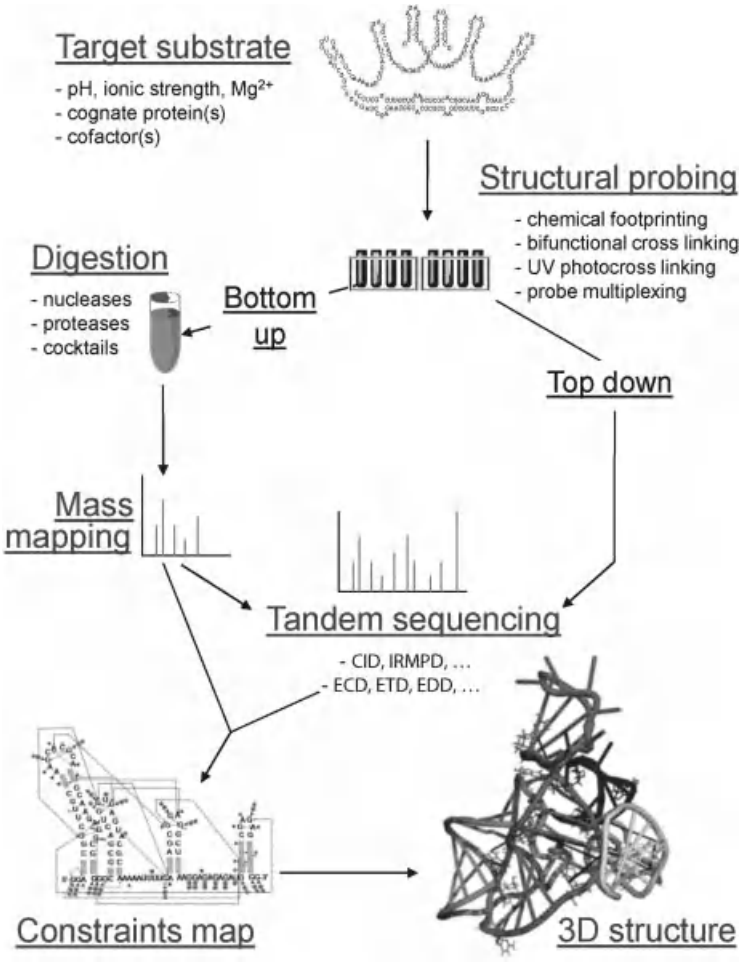


Figure 8.8. General workflow for three-dimensional (3D)-structure determination of nucleic acids based on structural probing and MS analysis (MS3D). The substrate is probed under ideal conditions preserving its native fold. Characterization of the ensuing covalent adducts can be performed under denaturing conditions, following either bottom-up or top-down approaches. The positions of probed nucleotides provide spatial constraints that are summarized on two-dimensional (2D) maps, from which a complete, all-atom 3D structure can be readily generated through established molecular-modeling protocols. [Reprinted with permission from (89). Copyright © 2010 American Society for Mass Spectrometry.] (See color version of the figure in Color Plate section)

these techniques, since protection against exchange upon binding is not always a direct measure of the binding site. Other structural changes, for instance, as a result of allosteric effects, may occur quite distant from the interaction surface, so caution is warranted when interpreting results of such studies. As with most scientific methods, a combination of techniques must be employed to obtain the most complete picture.

Of course, MS methods are also extremely powerful for measuring the stoichiometry of protein–DNA and protein–RNA interactions. The applications here seem limited only by imagination, with studies ranging from complexes of nucleic acids with small molecules, proteins (94,95), or other polynucleotides (47,96,97), right up to investigations of the components and structural organization of proteins and RNA in whole ribosomes. For example, the HIV SL-1 RNA dimer forms a complex with nucleocapsid protein that subsequently structurally isomerizes from a loop to an extended duplex (Fig. 8.9). These complexes have been detected directly by ESI MS, and then differentiated based on their different products formed under mild dissociation conditions (98). Larger protein–RNA complexes can also be structurally (or at least organizationally) elucidated based on a combination of altering solution conditions to disrupt native interactions, and measuring the order of dissociation of the complexes under mild conditions and the nature of the subassemblies detected in the mass spectra. This approach was applied recently to map the interactions within the human translation initiation factor eIF3 and how these are modulated by binding to the hepatitis C virus internal

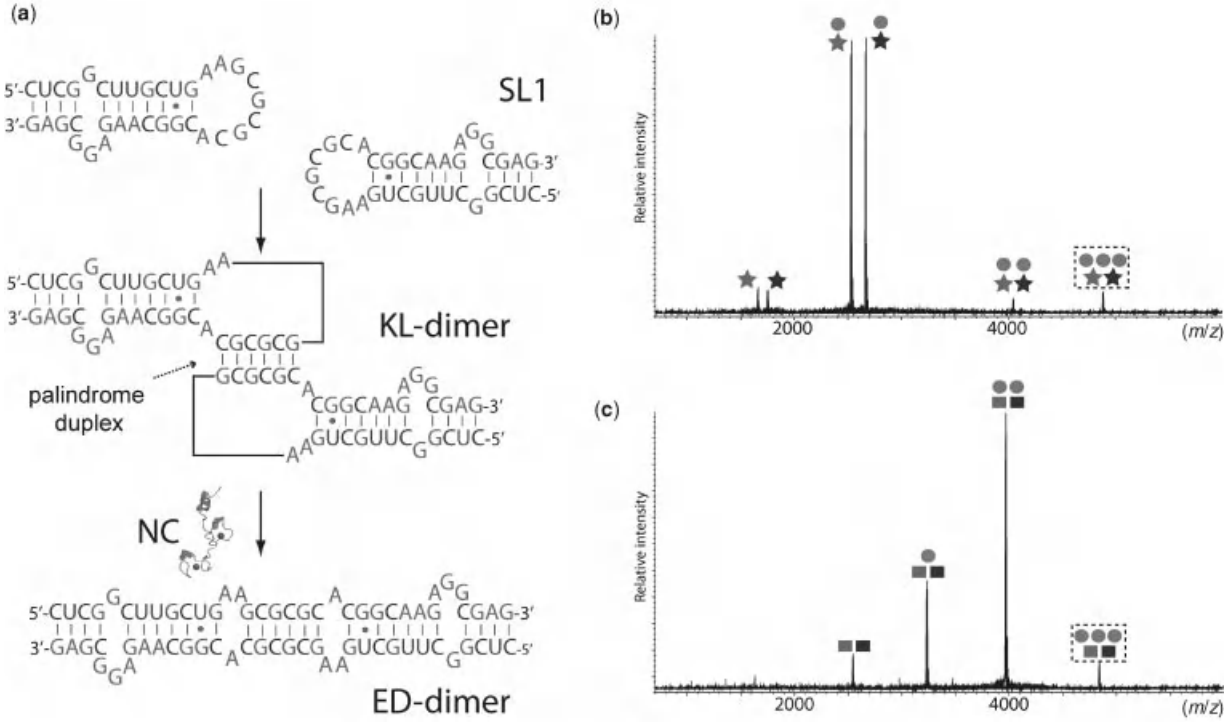


Figure 8.9. (*a*) Dimerization of HIV-1 SL1 into a kissing-loop (KL) complex and isomerization into an extended duplex (ED) mediated by nucleocapsid (NC) protein. Gentle CID of (*b*) NC•KL and (*c*) NC•ED dimer. Stars and squares identify KL- and ED-specific dimers. Circles identify NC. Precursor ions in dashed boxes possess identical masses. [Reprinted with permission from (36). Copyright © 2011 American Chemical Society.]

ribosome entry site (HCV–IRES) RNA (99). Further studies of large protein–RNA complexes are discussed in Chapter 9.

The kinetics of protein–RNA reactions can also be usefully monitored by MS. One example of this is the binding of ricin A to a short stem–loop hairpin segment of HIV Ψ-RNA. The former is a cytotoxic protein that hydrolyzes a specific N-glycosidic linkage, causing depurination at an adenine base in the sequence. The kinetics of depurination of the RNA molecule were measured by direct infusion ESI MS (100).

Mass spectrometry also clearly provides a relatively rapid screening method that can, at least somewhat quantitatively, assess ligand-binding affinities. Some of these techniques were already discussed elsewhere in this text (see Chapter 7 and Section 8.1.2). A popular screening method for RNA binders involves mixing a number of ligands simultaneously with the analyte of interest, from which we can obtain relative affinities in a competition-binding experiment. Only the strongest binders will form detectable complexes, and their relative intensity ratios give an indication of relative affinities (101,102). For instance, screening a large library of compounds against the HCV–IRES RNA mentioned above identified a new class of small molecule ligands that bind to the polynucleotide, and could potentially represent a therapeutic target (103). There is also evidence that, if sufficiently gentle source conditions are selected, and data analyzed rigorously, absolute dissociation constants can be obtained that are comparable to those measured by more traditional biophysical methods (104,105).

8.2. OLIGOSACCHARIDES

Until recently, the importance of carbohydrates has been somewhat overlooked relative to proteins, partly because their analysis is a significantly more complex task. Oligosaccharides can exist alone as polymers (e.g., cellulose or starch), but are also commonly found attached at various sites on the surface of proteins. In the wake of proteomics the relatively new field of *glycomics*, the study of protein glycosylation, has developed to understand the functional implications of carbohydrate production. Naturally occurring oligosaccharides attached as glycoconjugates to protein molecules generally consist of hexose sugar units (Fig. 8.10). These monomers can be linked to each other via any of their hydroxyl groups to form complex linear or branched polymers. Linkages are notated according to the numbering of the carbon atoms in the ring that constitute the cross-link, and the configuration at the anomeric carbon

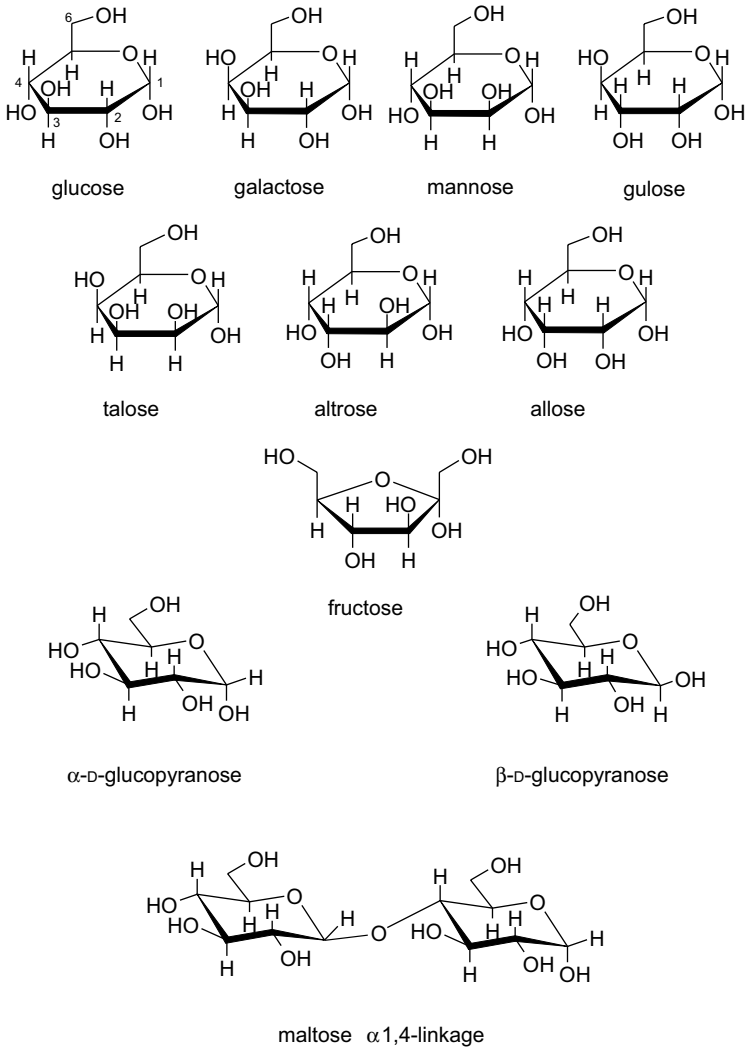


Figure 8.10. Structures of common saccharides. Most hexose sugars normally exist in the ribose form although fructose occurs naturally in the pyranose form. The α- and β- form of glucose differ by their orientation of the anomeric hydroxyl group. Linkages in oligosaccharides are designated by the anomeric configuration and the carbons linked at each monosaccharide unit.

(α or β). For instance, the disaccharide maltose consists of two D-glucose molecules linked via an $\alpha(1 \rightarrow 4)$ linkage (Fig. 8.10). Additionally, a number of other saccharide units occur naturally, most notably the pentose sugar ribose that forms the backbone of nucleic acids in genetic material.

The majority of proteins that are either secreted or found on the surface of cells become decorated with carbohydrates after translation. The exact purpose of these glycans is not yet fully understood, but they have been implicated in structure, folding, quality control, as signals for protein export, and as recognition sites for binding of other proteins in macromolecular assemblies. While the sites of glycosylation within a protein are generally highly specific, the nature of the carbohydrate is often less well defined. The primary amino acid sequence for a given protein is always the same, but post-translational modifications that add sugars to the surface of protein always lead to at least some polydispersity in the final molecule. This results from less tight control of the biosynthetic pathways for oligosaccharides. This may be a design feature allowing individual glycoprotein molecules to present similar, but slightly different, surface architectures for recognition. The variable nature of these sugars can lead, however, to obvious problems in characterization, since standard biochemical techniques of separation (e.g., chromatography or gel electrophoresis), result in broad peaks or bands from the inherent variability in molecular weight. Such polydispersity lends the subject of glycosylation to be studied by MS. It is therefore not surprising that MS has emerged as a key technique to obtain detailed structural information about the nature of the sugars and the mode of their attachment to proteins(106,107).

8.2.1. Covalent Structure of Oligosaccharides

The methodologies for analysis of proteins and peptides, at least in terms of sequence determination by tandem mass spectrometry (MS/MS), are now quite mature relative to the analysis of carbohydrate structure. As seen in Section 3.4, polypeptide fragmentation tends to follow well-defined pathways, primarily occurring along the peptide backbone, which conveniently generates fragments from which the sequence of the intact peptide can be derived (107). This is due to the inherently easier fission of peptide bonds by collisional activation relative to that of aliphatic side-chain linkages. By contrast, many of the bonds in carbohydrate molecules are chemically similar, leading to much more complex fragmentation pathways. Bond fission commonly occurs not only between saccharides, but also across the glycosidic ring. In addition, loss of water occurs quite trivially and multiple rearrangement pathways are available to activated species that render analysis of MS/MS data extremely complex. The nomenclature for designation of product ions produced by fragmentation of glycoconjugates is analogous to that used for peptide fragmentation and was developed by Domon and Costello (109), as shown in Figure 8.11.

With the use of ESI it is possible to obtain ionized molecules from underivatized oligosaccharides either as protonated, deprotonated, or alkali metal adduct ions. Under low-energy collision-induced dissociation (CID) conditions, the most common bond cleavages occur between the glycosidic rings, whereas high-energy dissociation can be employed to effect cross-ring cleavages, involving breaking two bonds, to yield further structural information. Differences in branching often can be detected due to the altered bond energies and steric effects that influence fragmentation pathways. Additionally, fragmentation is strongly influenced by the nature of the parent ion: alkali metal cationized species exhibit different fragmentation patterns from their protonated analogues, so a suite of different electrospray conditions can be employed to achieve more detailed data. Yet more information can be gained by using various derivatization techniques. One can also take advantage of the wealth of rearrangement reactions that occur and extrapolate back to the original structures that produced the observed product ions. However, if a parent ion produces a large number of fragments, then it can become difficult to obtain good ion statistics from the detection of product ions using scanning analyzers. Therefore, mass analyzers with high-duty cycles [e.g., TOF and the various trapping analyzers] have proved very important to investigate low-abundance species. Additionally, hybrid instruments with increased resolving power and mass accuracy for fragment ions will clearly be of great benefit.

Glycopeptides are another area of great interest, especially with the modern demands for high throughput analysis. The methods described above were developed for the analysis of purified oligosaccharides or mixtures of sugars. A more attractive approach in the realm of proteomics is to be able to derive the structural and sequence information of the sugar moiety attached *in situ* to its host protein. Thus, from a tryptic digest of a glycosylated protein, one could in theory determine not only the exact site of glycoside attachment on the protein under scrutiny, but also the nature of the oligosaccharide, in a single experiment. Glycosylation site analysis is typically carried out by identifying glycopeptides among proteolytic fragments (e.g., by comparing peptide maps for intact and deglycosylated protein). If peptide mapping of deglycosylated protein is not feasible (e.g., due to poor solubility of the carbohydrate-free form of the protein), glycopeptides can be identified in the digest of intact glycoprotein by observing characteristic losses (e.g., 162 Da for hexose residues) in survey MS/MS spectra obtained with CID of peptide ions, since the labile nature of glycosidic bonds in the gas phase leads to their facile dissociation. While this feature is certainly useful for identifying glycopeptides in complex mixtures, it may prevent more precise localization of glycosylation sites. In these cases, other fragmentation techniques [e.g., electron-capture dissociation (ECD) or electron-transfer dissociation (ETD)], can be effectively employed, as they preferentially cleave the peptide backbone, leaving the carbohydrate chains mostly intact (110).

If needed, glycan release from glycoproteins can be effected using peptide *N*-glycosidase F (PNGase F) and/or

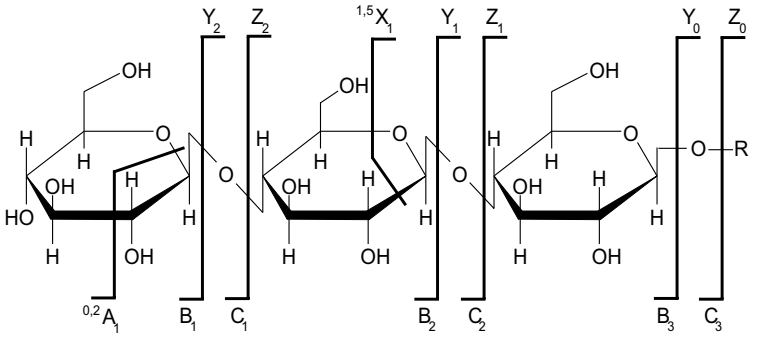


Figure 8.11. Fragmentation nomenclature for oligosaccharides of Domon and Costello (108).

PNGase A for *N*-glycans, and chemical methods (e.g., hydrazinolysis) for *O*-glycans (111). Methods of MS/MS, especially electron-based ion fragmentation methods (ECD and ETD), can be used in combination with exoglycosidases to obtain detailed structural information on released glycans (111–113). However, complete determination of structure (especially with novel glycans) frequently requires the use of orthogonal methods (e.g., NMR and X-ray crystallography) in addition to MS (111). The area of glycomics is a vast and rapidly developing field that we cannot hope to cover in sufficient detail in this text. We would point interested readers to recent excellent reviews of current methodology for oligosaccharide identification (113,114).

8.2.2. Higher Order Structure of Oligosaccharides and Interactions with their Physiological Partners

Carbohydrates are generally the most flexible type of biopolymers, and it is difficult to use the notion of a well-defined conformation, except for situations when a glycan chain is bound to a stable template, (e.g., its partner protein). In fact, the very term conformation is commonly used in carbohydrate chemistry in relation to configurations of individual sugar units, as it has been used for decades in organic chemistry. Still, the issue of carbohydrate higher order structure is now being explored at least for some glycan classes (115,116), keeping in mind the highly dynamic character of such architectures. While NMR and other spectroscopic methods appear to have been the major tools in these studies, MS is also making a debut, although it probably is still too early to say what impact it will make in this field.

As is the case with oligonucleotides, exchange of labile hydrogen atoms in carbohydrates occurs too quickly to be of any analytical utility [according to a recent report (116), one possible exception could be protons attached to acetamido groups]. Ironically, one of the first documented uses of HDX reactions in the biopolymer arena was in the analysis of carbohydrates (cellulose), where this technique was used as a means of counting the number of exchangeable hydrogen atoms to assist in elucidating the covalent structure of these macromolecules (117). Fast forward 70 years and HDX MS is still being used essentially as a hydrogen-counting tool for carbohydrate molecules (119), although some of the results of two recent studies may indicate that HDX MS can detect exchangeable hydrogen atoms that are either shielded from the solvent (119) or involved in hydrogen-bonding networks (117). Charge-state distribution analysis of ions produced by ESI, another powerful technique to assess conformation of proteins (see Section 5.1), has never been evaluated as a means to probe carbohydrate conformation, most likely due to the lack of proper standards with well-established solution geometry.

A few reports were published that document the use of IM–MS to study conformational characteristics of oligosaccharides in the gas phase. For example, Leary and co-workers (120) observed significant changes in collisional cross-section of short chains of a highly sulfated oligosaccharide heparin as a result of metal binding, and in some instances the presence of multiple conformations in the gas phase was evident. Although it is tempting to equate these structures to heparin conformation in solution, one must keep in mind that solvent removal may change the biopolymer structure, as discussed in Section 4.1. More to the point, the detailed study of oligonucleotide polyanions in the gas phase by Bowers and co-workers, (64) mentioned in Section 8.1.2) suggested that electrostatic repulsion in the gas phase leads to structure elongation. Again, it is difficult to say anything with certainty due to the absence of reliable standards (polysaccharides with well-defined conformations in solution) or computational tools that would allow the higher order structure to be predicted based on the chemical structure of these molecules.

There are many more examples of using ESI MS to characterize interaction of polysaccharides with other biomolecules, most notably proteins. This is typically done using direct ESI MS measurements, an approach discussed in detail in Section 7.1. Perhaps the most significant difference between the examples of protein–ligand binding discussed in Section 7.1 and protein–polysaccharide interactions is the notable heterogeneity of the pool of ligands in the latter case. Unless a well-defined synthetic polysaccharide is available for such studies, direct ESI MS measurements of protein–polysaccharide interactions have to deal with mixtures of ligands, which could exhibit a high degree of heterogeneity. Mass spectrometry often provides a means to identify binding-competent species in such very complex libraries. An example is presented in Figure 8.12, where low molecular weight polysaccharide heparin (Tinzaparin™, Tz) is mixed with a plasma protein antithrombin-III (AT-III), a cognate heparin binder. Although the heterogeneity of Tz is so great that it does not allow individual species to be resolved in the ESI MS, the data clearly indicate that only a few species from this vast pool bind to AT-III. Measurements similar to those shown in Figure 8.12 can also be used to rank polysaccharides according to their protein-binding affinities (121). Electrospray MS can also be used to determine absolute binding affinities of short oligosaccharides to their physiological protein targets, but this typically requires significantly less complex mixtures (122,123).

8.3. SYNTHETIC POLYMERS AND THEIR CONJUGATES WITH BIOMOLECULES

A common feature of the three classes of biopolymers considered so far is the unique sequence of the building

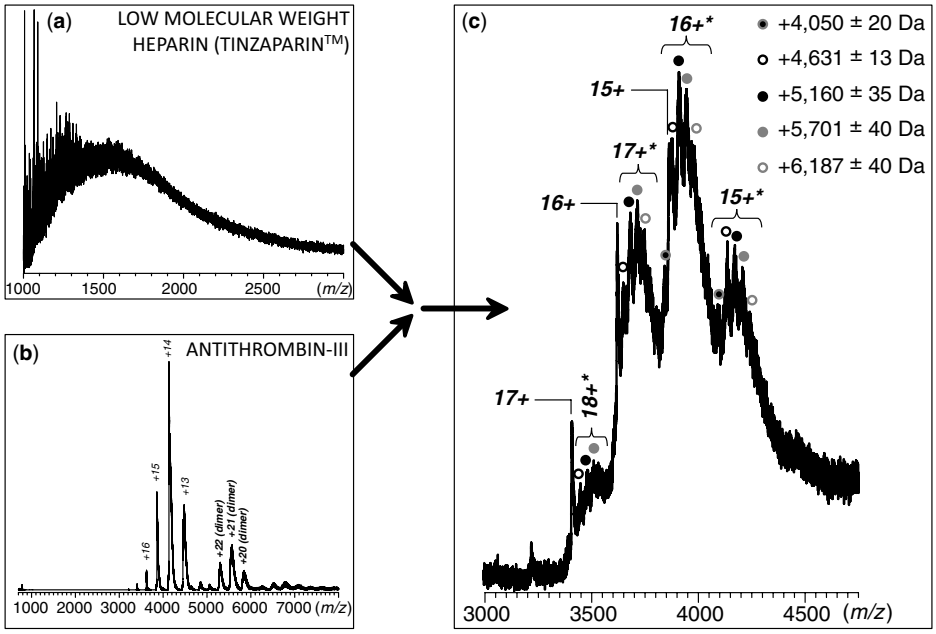


Figure 8.12. Electrospray MS characterization of binding low molecular weight heparin chains to AT-III. The mass spectra of Tinzaparin™ (*a*) and free AT-III (*b*). The spectrum in (*c*) shows that only a few select heparinoids bind to the protein. [Adapted with permission from (120). Copyright © 2007 American Chemical Society.]

blocks (e.g., amino acid sequence of proteins) that defines their structure, dynamic behavior, and functional properties. However, there is also a variety of biopolymers that do not display biological activity on their own, but instead play *passive* (mostly structural) roles in living organisms. Structural integrity and cohesion (from the organelle to the organism level) are vital for all living creatures, hence the importance of passive biopolymers. Since the function of such biopolymers is usually limited to structural reinforcement, they often display significant sequence redundancy (Fig. 8.13). Perhaps the most illustrious example of structural proteins is fibroin, a fibrous protein that makes up most silks. For example, the heavy chain of *Bombyx mori* silk fibroin consists of 5263 amino acid residues, one-half of which are glycine and one-third are alanine residues (124). The sequence of this protein contains 433 repeat units GAGAGS that have high β-sheet propensity and form the structural basis for the remarkable physical and mechanical properties of silk (125,126). Another example of a structural protein is collagen (127), a family of large proteins (>1000 amino acid residues per single chain) that make up almost a third of the total protein mass in multicellular organisms. The fibril-forming types of collagen are rich in GPX repeat units (Fig. 8.13), a structural element that forces collagen polypeptide chains to adopt an unusual left-handed helical conformation, three of which twist together to form a right-handed super-helical structure (128).

Other examples of passive biopolymers are presented by various homopolysaccharides, most of which are structural polymers, while others are energy-storing polymers. Cellulose (Fig. 8.13) is a typical example of a structural polysaccharide and is the most abundant organic substance on the planet (it is estimated that $\sim 10^{11}$ tons of cellulose is produced photosynthetically each year). A single cellulose molecule can contain as many as 15,000 D-glucose residues linked by $\beta(1 \rightarrow 4)$ glycosidic bonds (like most other polysaccharides discussed in Section 8.2, cellulose does not have a well-defined size). The exceptional strength of cellulose fibers is due to the formation of multiple strong inter- and intramolecular hydrogen bonds within the cellulose fibers (129,130). Additional structural reinforcement of cellulose fibers in wood is provided by a cementing matrix, whose major component is lignin, a plastic-like phenolic polymer (Fig. 8.13).

The examples of biopolymers considered above represent several major classes of polymer molecules. When a polymer chain is made by linking only one type of repeat unit, (e.g., D-glucose residue in cellulose), it is called a *homopolymer*. Natural rubber, or polyisoprene (Fig. 8.13), is another example of a homopolymer. When two (or more) different types of repeat units are joined in the same polymer chain, the polymer is called a *copolymer*. Fibroin and collagen fall into this category, as they contain several building blocks (amino acid residues). Unlike fibroin and cellulose, lignin structure is not linear, but rather represents a *random 3D network polymer*.

All of the aforementioned polymers are naturally occurring and of biological origin. Of great interest is the possibility of producing polymeric materials synthetically

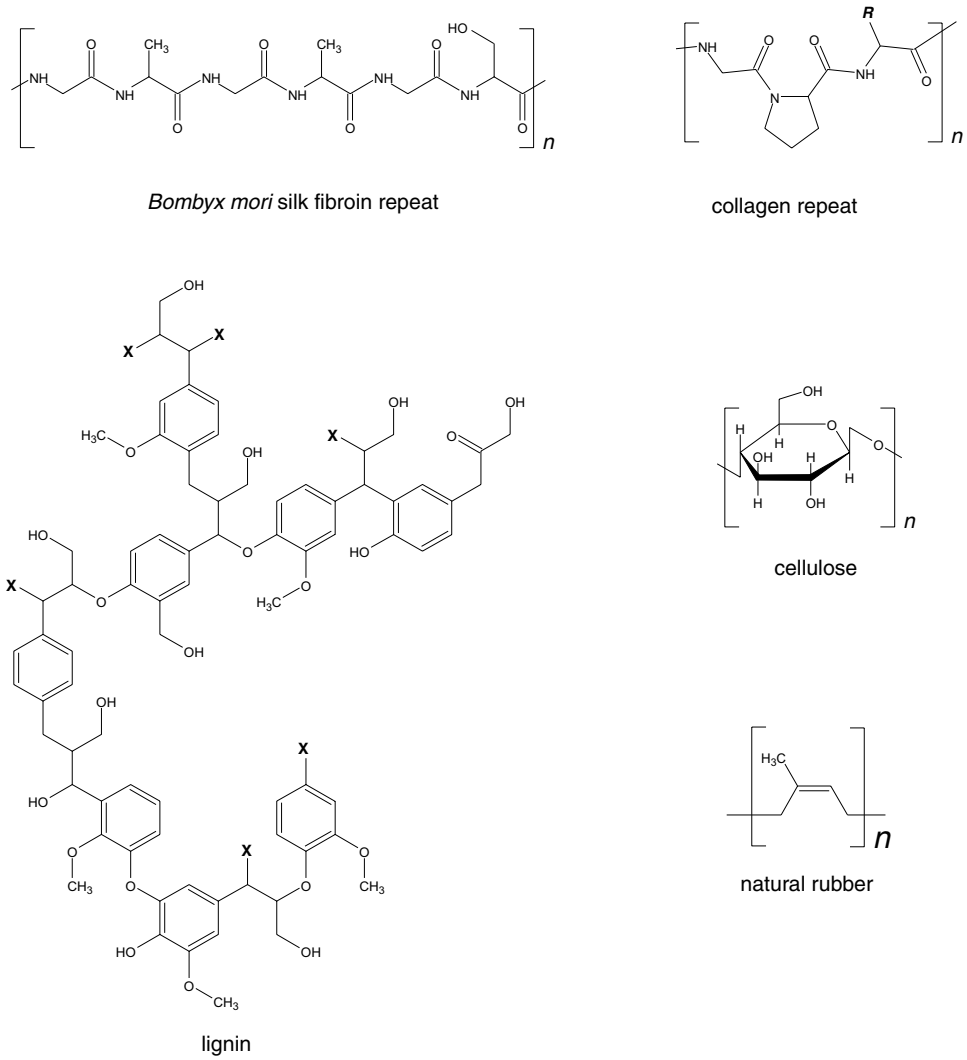


Figure 8.13. Examples of structural biopolymers of biological origin.

that either mimic their biological analogues, or indeed improve on their properties in some way. Following the pioneering work by Staudinger (131,132) and Carothers et al. (133), the macromolecular nature of synthetic polymers has become a commonly accepted concept. Although a detailed consideration of synthetic polymers is beyond the scope of this book, we will briefly discuss several classes of polymers whose behavior is important from the biophysical point of view. The first class is a rapidly growing group of the so-called "bioinspired" polymers that are designed to possess certain structural or functional features of biological macromolecules (134–136). Fibroin, mentioned earlier, has been a particularly popular target for such design due to its superior mechanical properties (125), with nylon (Fig. 8.14) representing perhaps the first successful attempt to imitate a biopolymer.

While the design of synthetic analogues of silk proteins remain the focal point of extensive research efforts (137), strategies that employ genetic engineering to mix the modules of the natural protein in specific proportions to attain the desired properties are rapidly gaining popularity (139). An important issue here is how the mechanical properties of the polymers are modulated by conformational transitions. An example of such conformational change is a helix-to-sheet transition in the amorphous segments of the dragline silk proteins, which is often invoked to explain its superior expandability characteristics. Understanding the general rules of construction of silk proteins, and the analysis of the relationship between their conformational and mechanical properties, will undoubtedly provide a guide to achieving the desired mechanical properties in manmade materials by designing controllable conformational switches (139).

Another class of synthetic polymers whose behavior in solution is beginning to attract the attention of biophysicists is comprised of several hydrophilic and amphiphilic macromolecules that are utilized in the design and therapeutic

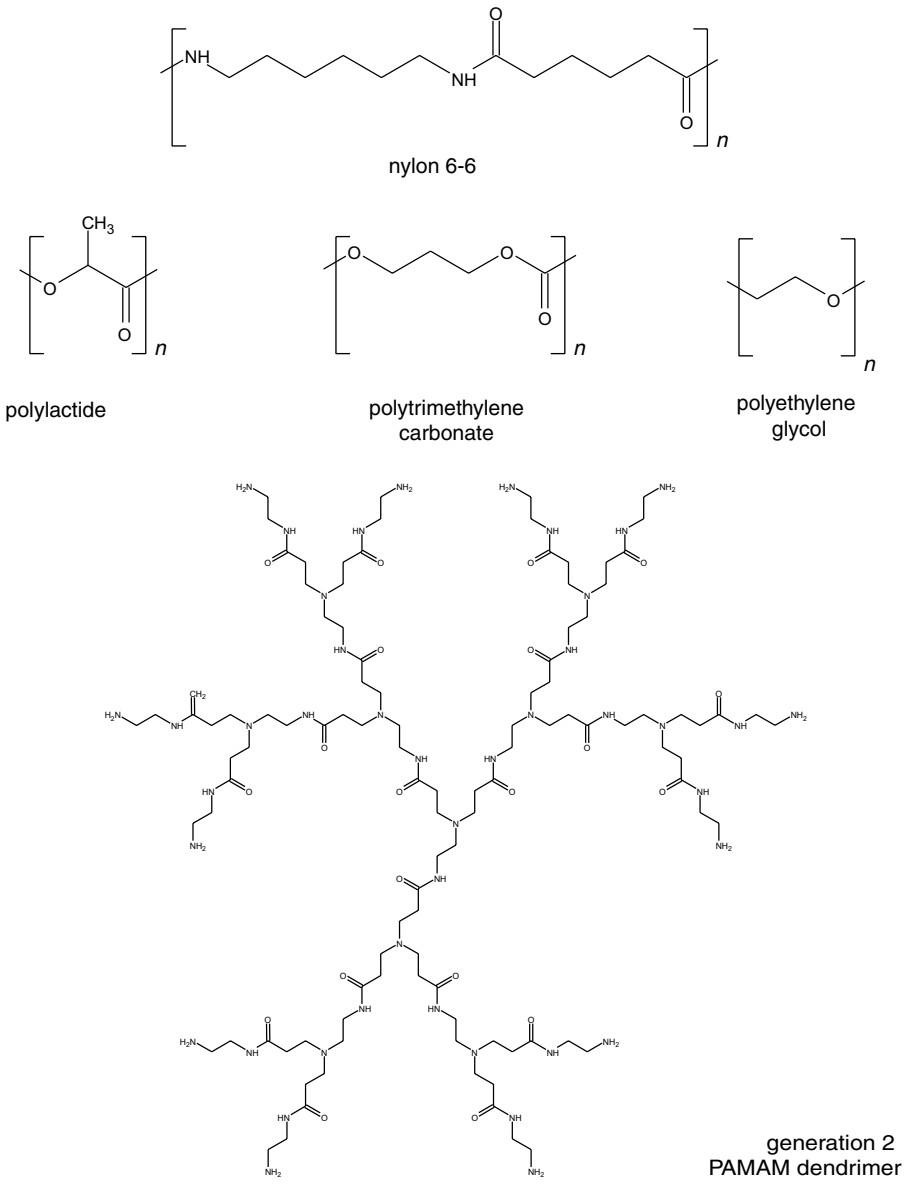


Figure 8.14. Examples of synthetic polymers. Polyamidoamine = PAMAM.

utilization of polymer-conjugated proteins (140–142). Conjugation of therapeutically active proteins with polyethylene glycol (PEG, Fig. 8.14) or other polymers improves their solubility, extends lifetime, and often modulates release. In the past, it was almost implicitly assumed that the *polymer tail* is fully unstructured in solution, and does not interact with the protein. This notion, however, is now being increasingly scrutinized. Indeed, polymer chains can collapse under certain conditions. Furthermore, electrostatic interaction can exist between the polymer tail and the protein, potentially interfering with the function of the latter. Clearly, the ability to monitor both large-scale dynamics and interaction in this system will provide valuable insight into behavior of the PEG-conjugated proteins.

8.3.1. Covalent Structure of Polymers and Polymer–Protein Conjugates

One striking difference between synthetic polymers and biopolymers whose production is genetically controlled (i.e., proteins and oligonucleotides) is the intrinsic heterogeneity of the former. To illustrate this point, we will consider the MALDI mass spectra of a 5.5-kDa polypeptide insulin and a synthetic polymer of similar mass, PEG-5000 (Fig. 8.15). Clearly, even the *monodisperse* PEG-5000 sample actually consists of over a dozen oligomers of different length. Such heterogeneity presents a significant challenge for polymer analysis, particularly when ESI MS is used, as the total ion signal is divided not only among

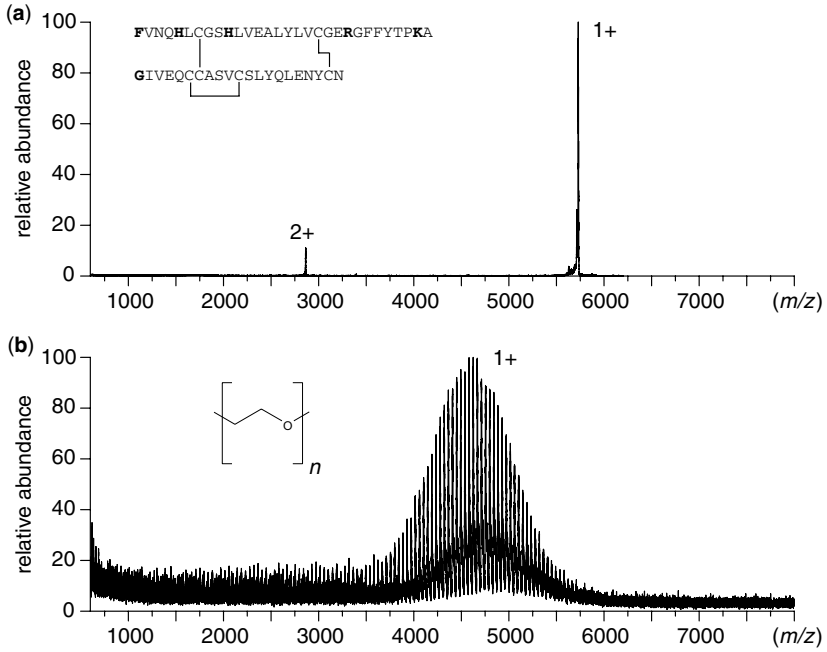


Figure 8.15. The MALDI mass spectra of a 5.5-kDa polypeptide insulin (*a*) and a monodisperse synthetic polymer PEG 5000 (*b*).

different charge states(as is the case with essentially monodisperse proteins), but also among chains differing in the number of blocks (degree of polymerization).

The structural heterogeneity of protein–polymer conjugates is even higher, as it exists at three different levels. Indeed, various protein isoforms may differ from each other by (1) the number of PEG chains attached to a single polypeptide chain, (2) location of the conjugation sites, and (3) the lengths of PEG chains. Protein separation methods, particularly size exclusion chromatography (SEC), in most cases can readily resolve isoforms of the first type based on significant differences in their physical size in solution. The extent of heterogeneity associated with polydispersity of PEG chain length can be visualized and evaluated by MALDI MS, as shown in Figure 8.16 for PEGylated ubiquitin. This protein has eight primary amines (seven lysine residues and the N-terminus), and its conjugation with activated 5-kDa PEG using reductive alkylation

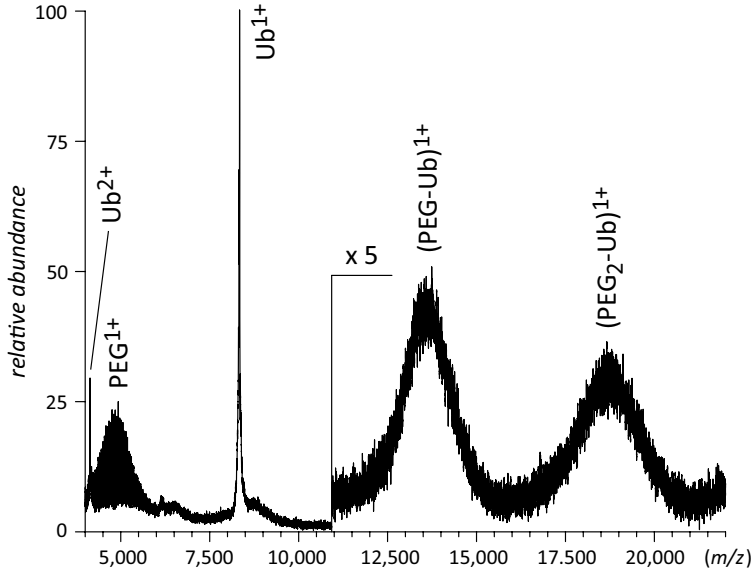


Figure 8.16. The MALDI TOF mass spectrum of unseparated products of ubiquitin PEGylation. [Reprinted Copyright © 2011 from (143) with permission from Elsevier Ltd.]

chemistry generates a range of products as revealed by MALDI MS analysis. However, it is the heterogeneity of the second type (protein–polymer conjugates differing in position of PEGylation sites in the polypeptide sequence) that is usually most difficult to evaluate. Its extent among the products of ubiquitin PEGylation can be appreciated by considering the ion exchange chromatogram (IXC) of the conjugation reaction products, which reveals at least five major peaks besides the peak representing unconjugated ubiquitin (Fig. 8.17*a*).

Peaks labeled 1–3 in Figure 8.17 all represent mono-PEGylated products (see Fig. 8.17*c*), and differ from one another by the location of the site of conjugation within the protein sequence. Identification of conjugation sites can be accomplished by comparing the peptide map of a particular form of the conjugated protein to that of the unmodified protein (144–146). Recently, feasibility of the top-down approach to localization of PEGylation sites was also demonstrated (143,147,148), an approach that relies on fragmenting the protein–polymer conjugate in the gas phase to obtain structural information. Gas-phase fragmentation of such species is a convoluted process, as it involves dissociation of covalent bonds in the polypeptide backbone, as well as the polymer chain. Fortunately, collisional activation of the PEG chain in particular leads to its efficient fragmentation, leaving the polypeptide chain intact (143). This generates protein ions with short truncated oligoethylene glycol units (Fig. 8.18), which can be used as placeholders, allowing the conjugation site to be localized in the second-stage fragmentation (143). Another fortunate consequence of PEG chain truncation is significant reduction of the heterogeneity of the ionic population, which allows direct

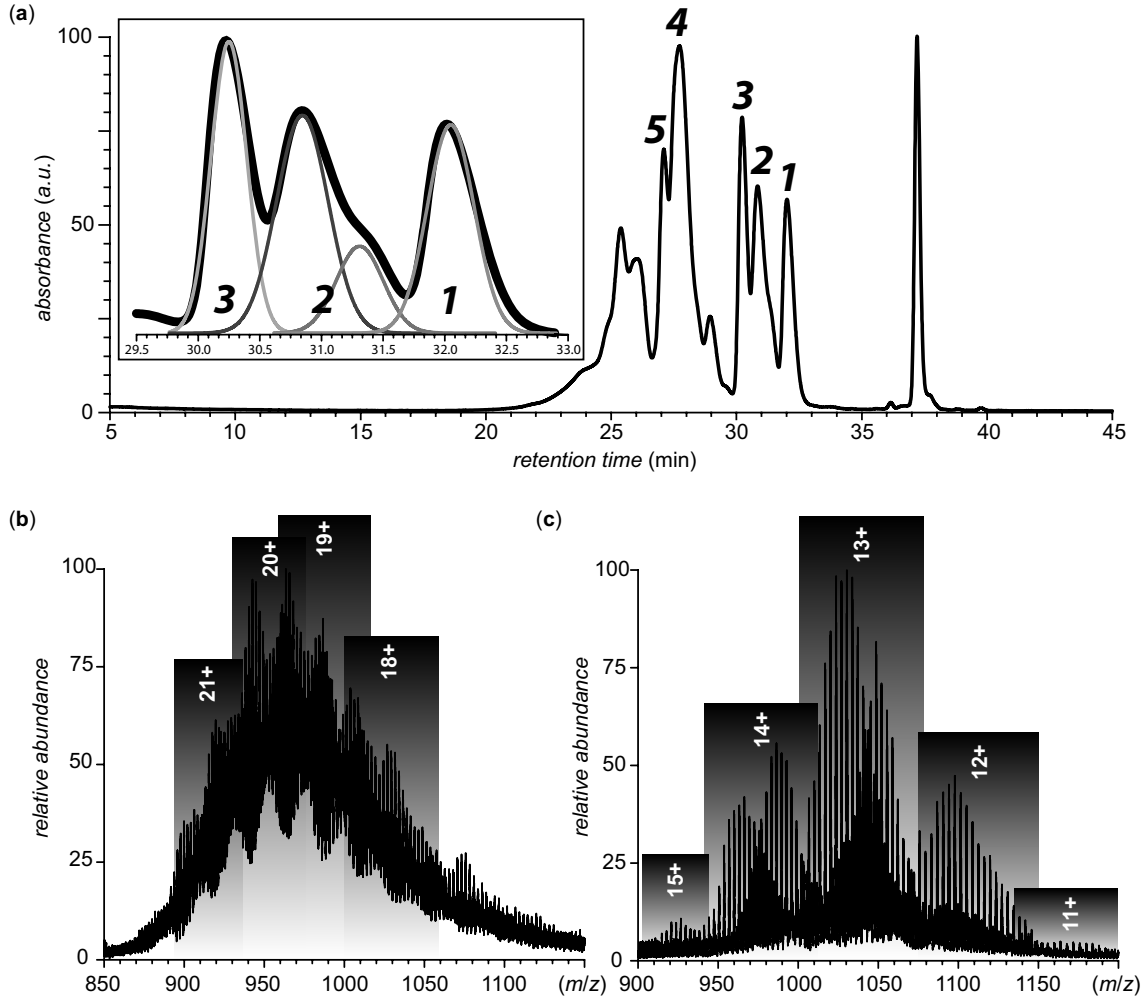


Figure 8.17. The IXC profile of PEGylated ubiquitin (*a*) and ESI mass spectra of pooled fractions representing IXC peaks 4 and 5 (*b*) and 1, 2, and 3 (*c*). The magnified view of the chromatogram (inset in a) shows the elution profiles of mono-PEGylated product (black trace), and deconvoluted contributions of individual isomers (gray traces). [Adapted Copyright © 2011 from (143) with permission from Elsevier Ltd.]

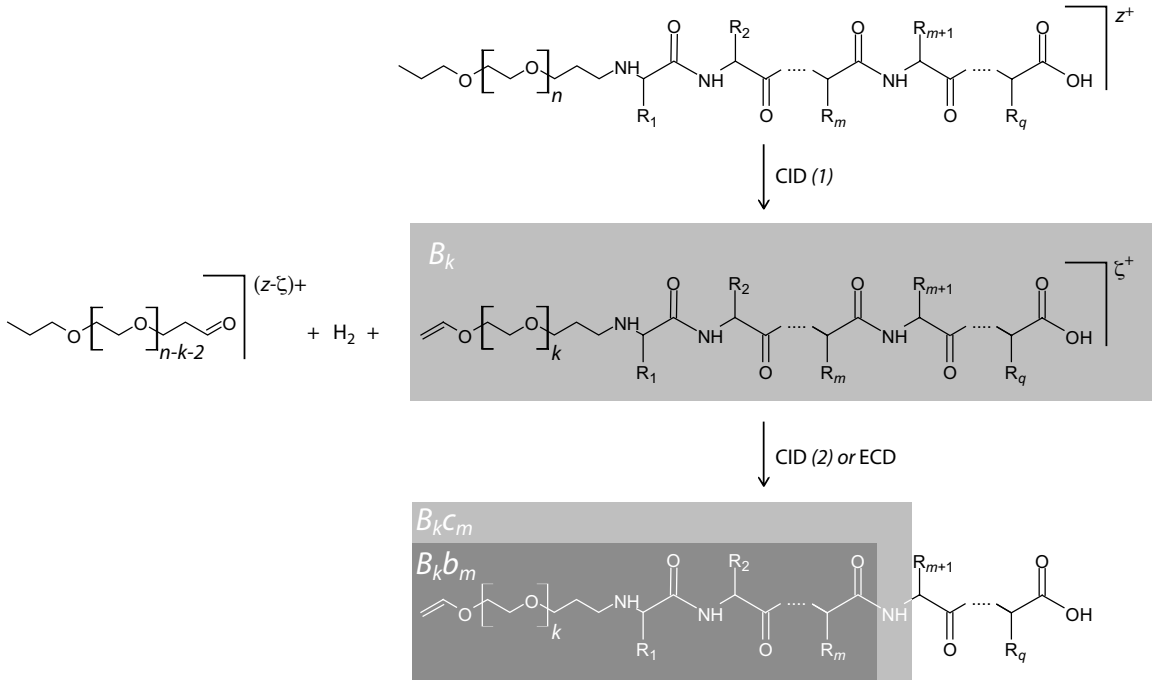


Figure 8.18. Proposed nomenclature of fragment ions derived from PEGylated proteins. [Reprinted Copyright © 2011 from (143) with permission Elsevier Ltd.]

measurements of global HDX MS to be carried out, as we will see in Section 8.3.2.

8.3.2. Higher Order Structure of Polymers and Polymer–Protein Conjugates

Among the many MS-based techniques considered in the previous chapters, analysis of the ionic charge-state distributions appears to be particularly promising as far as the analysis of conformational dynamics of water-soluble synthetic polymers and their conjugates with proteins. Although the heterogeneity of such species undoubtedly presents a challenge for the analysis of ionic charge-state distributions in ESI MS, utilization of gas-phase ion chemistry in combination with mathematical tools may allow this problem to be overcome. In Chapter 9, we will see an example of how this approach can be used to work even with highly heterogeneous systems, where MS fails to resolve individual charge states.

Another approach that we have found to be useful as a means of characterizing higher order structure and dynamics of PEGylated proteins is HDX MS. Although application of this technique at the whole conjugate level is problematic, due to the heterogeneity of these species, truncation of the PEG chain results in significant reduction of polydispersity of the ionic population (143). As a result, it becomes possible to measure the rates of global exchange within the PEGylated protein (Fig. 8.19). Feasibility of local (bottom-up) HDX MS measurements in PEGylated proteins was also recently demonstrated (149).

A variety of synthetic polymers are known to have stable secondary structures in solution (134,150), which are often maintained by elaborate networks of hydrogen bonds. Although in principle such networks can be characterized by measuring the kinetics of hydrogen exchange (similar to protein HDX discussed in previous chapters), there are very few studies that actually employ this technique as a tool to probe higher order structure of synthetic polymers (151). This is certainly an area where MS may make a significant contribution in the near future. One class of synthetic polymers that appears to be particularly suited for HDX studies is amide-based dendrimers [e.g., PAMAM], presented in Figure 8.14. Although the dendrimeric structure is usually presented in a highly symmetrical fashion, these branched polymers may actually exhibit a variety of conformations. Since the dendrimer architecture is an important determinant of its physical properties, numerous studies have aimed at elucidating the structure of dendrimers (particularly, amphiphilic dendrimers) in various environments, often revealing quite unexpected conformational features of these macromolecules (152). At high ionic strength, backfolding of the end groups takes place, leading to formation of a *dense core* structure, while low ionic strength forces the dendrimer to stretch and adopt a *dense shell* structure (151). The HDX rates of some alanine-based dendrimers have been measured by ^{1}H NMR spectroscopy in polar solvents favoring the dense shell

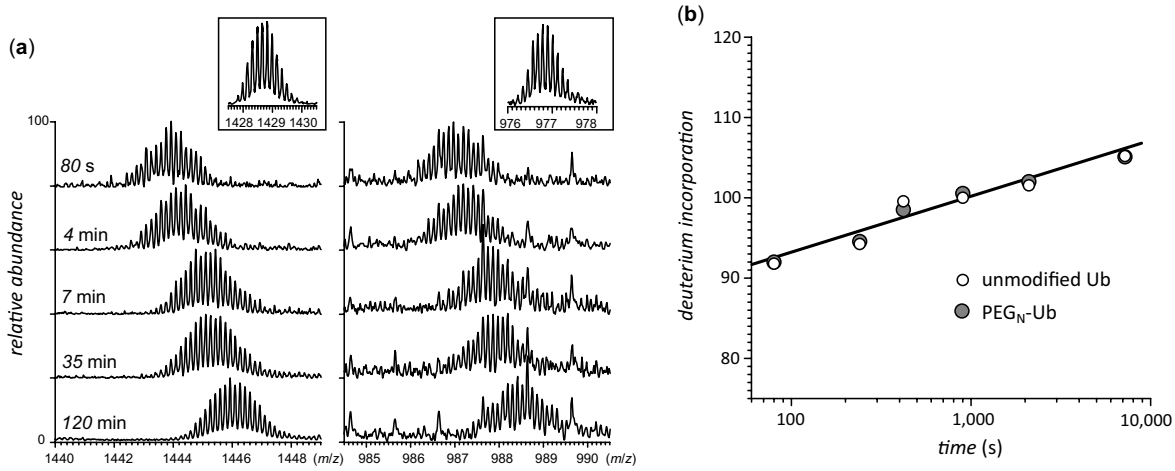


Figure 8.19. (*a*) Global hydrogen exchange in unmodified ubiquitin (left) and PEGylated ubiquitin (right) measured as a mass increase following protein exposure to D_2O based buffer for varying periods of time as indicated on the panels. The PEGylated ubiquitin ions were collisionally activated in the gas phase prior to mass measurement to truncate the polymer chain and reduce the heterogeneity of the ionic ensemble. Insets in this panel show reference mass spectra of unexchanged ions of ubiquitin and PEG–ubiquitin. (*b*) shows the exchange kinetics extracted from the HDX MS data shown in (*a*).

conformation (153). The dense core–dense shell transition certainly changes solvent accessibility of the amide groups, which may, in principle, be detectable by HDX MS. However, these structures are likely to be highly dynamic, presenting a formidable challenge to HDX MS measurements in solution, as is the case with carbohydrates and oligonucleotides. Again, gas phase HDX may present an attractive, but as yet unexplored, alternative to solution-phase HDX in this case, with the added benefit of allowing the structural heterogeneity to be greatly reduced prior to the exchange reactions using the ionic mass selection capability of MS.

Another approach to the structural characterization of dendrimers involves using MS/MS techniques. A combination of MALDI and ESI MS/MS techniques was previously employed to investigate the fragmentation pathways of a poly(propylene imine) dendrimer(154). Depending on the polarity of the solvent, different dissociation patterns were observed: In polar protic solvents an extended conformation was adopted due to favorable solvent–solute interactions. In this case, terminal branches were readily cleaved in gas-phase dissociation events, whereas the compact structure with intramolecular hydrogen bonding adopted in nonpolar solvents precluded such fragmentations.

Others, such as PAMAM dendrimers, showed significant potential as efficient nonviral vehicles for delivering genetic material into cells. They were shown to be as efficient or more efficient than either cationic liposomes or other cationic polymers (e.g., polyethylenimine or polylysine) for *in vitro* gene transfer (155). Mass spectrometry may be of use in determining how these molecules package genetic material and enable its introduction into cells, potentially directing further synthetic strategies.

As we have seen, the full power of MS in the non-protein biological arena has yet to be wholly exploited. Techniques that already exist, and those that are still under development, show great promise for the analysis of structure and dynamics in oligosaccharides, oligonucleotides, and even synthetic polymers.

REFERENCES

1. Jacobson, K. B., Arlinghaus, H. F., Buchanan, M. V., Chen, C. H., Glish, G. L., Hettich, R. L., and McLuckey, S. A. (1991) Applications of mass spectrometry to DNA sequencing, *Gen. Anal.-Biomo. Eng. 8*, 223–229.
2. Smith, L. M. (1993) The future of DNA sequencing, *Science 262*, 530–532.
3. Murray, K. K. (1996) DNA sequencing by mass spectrometry, *J. Mass Spectrom. 31*, 1203–1215.
4. Smith, L. M. (1996) Sequence from spectrometry: A realistic prospect? *Nat. Biotechnol. 14*, 1084–1088.
5. Spengler, B., Pan, Y., Cotter, R. J., and Kan, L. S. (1990) Molecular weight determination of underivatized oligodeoxyribonucleotides by positive ion matrix assisted ultraviolet laser desorption mass spectrometry, *Rapid Commun. Mass Spectrom. 4*, 99–102.
6. Covey, T. R., Bonner, R. F., Shushan, B. I., and Henion, J. (1988) The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry, *Rapid Commun. Mass Spectrom. 2*, 249–256.

7. Maxam, A. M. and Gilbert, W. (1977) A new method for sequencing DNA, *Proc. Natl. Acad. Sci. USA 74*, 560–564.
8. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA 74*, 5463–5467.
9. McLuckey, S. A., Van Berkel, G. J., and Glish, G. L. (1992) Tandem mass spectrometry of small, multiply charged oligonucleotides, *J. Am. Soc. Mass Spectrom. 3*, 60–70.
10. Little, D. P., Thannhauser, T. W., and McLafferty, F. W. (1995) Verification of 50-mer to 100-mer DNA and RNA sequences with high resolution mass spectrometry, *Proc. Natl. Acad. Sci. USA 92*, 2318–2322.
11. McLuckey, S. A. and Stephenson, *J. _L., Jr.* (1998) Ion/ion chemistry of high-mass multiply charged ions, *Mass Spectrom. Rev. 17*, 369–407.
12. Turner, K. B., Monti, S. A., and Fabris, D. (2008) Like polarity ion/ion reactions enable the investigation of specific metal interactions in nucleic acids and their noncovalent assemblies, *J. Am. Chem. Soc. 130*, 13353–13363.
13. Zhu, L., Parr, G. R., Fitzgerald, M. C., Nelson, C. M., and Smith, L. M. (1995) Oligodeoxynucleotide fragmentation in MALDI/TOF mass spectrometry using 355 nm radiation, *J. Am. Chem. Soc. 117*, 6048–6056.
14. Nordhoff, E., Karas, M., Cramer, R., Hahner, S., Hillenkamp, F., Kirpekar, F., Lezius, A., Muth, J., Meier, C., and Engels, J. W. (1995) Direct mass spectrometric sequencing of low picomole amounts of oligodeoxynucleotides with up to 21 bases by matrix assisted laser desorption ionization mass spectrometry, *J. Mass Spectrom. 30*, 99–112.
15. Ono, T., Scalf, M., and Smith, L. M. (1997) 2'-fluoro modified nucleic acids: polymerase-directed synthesis, properties and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry, *Nucleic Acids Res. 25*, 4581–4588.
16. Schneider, K. and Chait, B. T. (1995) Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix assisted laser desorption mass spectrometry, *Nucleic Acids Res. 23*, 1570–1575.
17. Tang, W., Zhu, L., and Smith, L. M. (1997) Controlling DNA fragmentation in MALDI-MS by chemical modification, *Anal. Chem. 69*, 302–312.
18. Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Hahner, S., and Hillenkamp, F. (1995) 7-deaza purine bases offer a higher ion stability in the analysis of DNA by matrix assisted laser desorption/ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 9*, 525–531.
19. Köster, H., Tang, K., Fu, D. J., Braun, A., vandenBoom, D., Smith, C. L., Cotter, R. J., and Cantor, C. R. (1996) A strategy for rapid and efficient DNA sequencing by mass spectrometry, *Nat. Biotechnol. 14*, 1123–1128.
20. Schieltz, D. M., Chou, C. W., Luo, C. W., Thomas, R. M., and Williams, P. (1992) Mass spectrometry of DNA mixtures by laser ablation from frozen aqueous solution, *Rapid Commun. Mass Spectrom. 6*, 631–636.
21. Williams, P. (1994) Time of flight mass spectrometry of DNA laser ablated from frozen aqueous solutions—applications to the Human Genome Project, *International Journal of Mass Spectrometry and Ion Processes 131*, 335–344.
22. Sheffer, J. D. and Murray, K. K. (2000) Infrared matrix-assisted laser desorption/ionization using a frozen alcohol matrix, *J. Mass Spectrom. 35*, 95–97.
23. Chen, X. Y., Westphall, M. S., and Smith, L. M. (2003) Mass spectrometric analysis of DNA mixtures: Instrumental effects responsible for decreased sensitivity with increasing mass, *Anal. Chem. 75*, 5944–5952.
24. Cheng, X. H., Camp, D. G., Wu, Q. Y., Bakhtiar, R., Springer, D. L., Morris, B. J., Bruce, J. E., Anderson, G. A., Edmonds, C. G., and Smith, R. D. (1996) Molecular weight determination of plasmid DNA using electrospray ionization mass spectrometry, *Nucleic Acids Res. 24*, 2183–2189.
25. Schafer, A. J. and Hawkins, J. R. (1998) DNA variations and the future of human genetics, *Nat. Biotechnol. 16*, 33–39.
26. Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., and Lander, E. S. (1998) Large scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome, *Science 280*, 1077–1082.
27. Brookes, A. J. (1999) The essence of SNPs, *Gene 234*, 177–186.
28. Henry, C. (1997) Can MS really compete in the DNA world?, *Anal. Chem. 69*, A243–A246.
29. Griffin, T. J., and Smith, L. M. (2000) Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry, *Trends Biotechnol. 18*, 77–84.
30. Tost, J. and Gut, I. G. (2005) Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications, *Clin. Biochem. 38*, 335–350.
31. Faulstich, K., Worner, K., Brill, H., and Engels, J. W. (1997) A sequencing method for RNA oligonucleotides based on mass spectrometry, *Anal. Chem. 69*, 4349–4353.
32. Little, D. P., Aaserud, D. J., Valaskovic, G. A., and McLafferty, F. W. (1996) Sequence information from 42-108-mer DNAs (complete for a 50-mer) by tandem mass spectrometry, *J. Am. Chem. Soc. 118*, 9352–9359.
33. Hofstadler, S. A., Griffey, R. H., Pasa-Tolic, L., and Smith, R. D. (1998) The use of a stable internal mass standard for accurate mass measurements of oligonucleotide fragment ions using electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry with infrared multiphoton dissociation, *Rapid Commun. Mass Spectrom. 12*, 1400–1404.
34. Limbach, P. A., Crain, P. F., and Mccloskey, J. A. (1994) The modified nucleosides of RNA—summary, *Nucleic Acids Res. 22*, 2183–2196.
35. Mcluckey, S. A. and Habibigoudarzi, S. (1993) Decompositions of multiply-charged oligonucleotide anions, *J. Am. Chem. Soc. 115*, 12085–12095.
36. Fabris, D. (2011) MS analysis of nucleic acids in the post-genomic era, *Anal. Chem. 83*, 5810–5816.

37. Kubicek, S. (2011) Epigenetics: A primer, *Scientist 25*, 32–33.
38. Li, G. and Reinberg, D. (2011) Chromatin higher-order structures and gene regulation, *Curr. Opin. Genet. Dev. 21*, 175–186.
39. Englander, S. W., Sosnick, T. R., Englander, J. J., and Mayne, L. (1996) Mechanisms and uses of hydrogen exchange, *Curr. Opin. Struct. Biol. 6*, 18–23.
40. Black, B. E., Foltz, D. R., Chakravarthy, S., Luger, K., Woods, V. L., and Cleveland, D. W. (2004) Structural determinants for generating centromeric chromatin, *Nature (London) 430*, 578–582.
41. Hansen, J. C., Wexler, B. B., Rogers, D. J., Hite, K. C., Panchenko, T., Ajith, S., and Black, B. E. (2011) DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2), *J. Biol. Chem. 286*, 18938–18948.
42. Suhnel, J. (2001) Beyond nucleic acid base pairs: from triads to heptads, *Biopolymers 61*, 32–51.
43. Balasubramanian, S., Hurley, L. H., and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?, *Nat. Rev. Drug Discov. 10*, 261–275.
44. Krishnan, Y. and Simmel, F. C. (2011) Nucleic acid based molecular devices, *Angew. Chem. Int. Ed. Endg. 50*, 3124–3156.
45. Han, F. X. G., Wheelhouse, R. T., and Hurley, L. H. (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition, *J. Am. Chem. Soc. 121*, 3561–3570.
46. Lightwahl, K. J., Springer, D. L., Winger, B. E., Edmonds, C. G., Camp, D. G., Thrall, B. D., and Smith, R D., (1993) Observation of a small oligonucleotide duplex by electrospray ionization mass-spectrometry, J. Am. Chem. Soc. *115*, 803–804.
47. Goodlett, D. R., Camp, D. G., Hardin, C. C., Corregan, M., and Smith, R. D. (1993) Direct observation of a DNA quadruplex by electrospray ionization mass spectrometry, *Biol. Mass Spectrom. 22*, 181–183.
48. Smargiasso, N., Rosu, F., Hsia, W., Colson, P., Baker, E. S., Bowers, M. T., De Pauw, E., and Gabelica, V. (2008) G-quadruplex DNA assemblies: Loop length, cation identity, and multimer formation, *J. Am. Chem. Soc. 130*, 10208–10216.
49. Rosu, F., Gabelica, V., Poncelet, H., and De Pauw, E. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry, *Nucleic Acids Res. 38*, 5217–5225.
50. Gale, D. C., Goodlett, D. R., Lightwahl, K. J., and Smith, R. D. (1994) Observation of duplex DNA-drug noncovalent complexes by electrospray-ionization mass-spectrometry, *J. Am. Chem. Soc. 116*, 6027–6028.
51. Rosu, F., De Pauw, E., and Gabelica, V. (2008) Electrospray mass spectrometry to study drug-nucleic acids interactions, *Biochimie 90*, 1074–1087.
52. Brodbelt, J. S. (2010) Evaluation of DNA/ligand interactions by electrospray ionization mass spectrometry, *Annu. Rev. Anal. Chem. 3* 67–87.
53. Lombardo, C. M., Martinez, I. S., Haider, S., Gabelica, V., De Pauw, E., Moses, J. E., and Neidle, S. (2011) Structure-based design of selective high-affinity telomeric quadruplex-binding ligands, *Chem. Commun. 46*, 9116–9118.
54. Guo, X., Bruist, M. F., Davis, D. L., and Bentzley, C. M. (2005) Secondary structural characterization of oligonucleotide strands using electrospray ionization mass spectrometry, *Nucl. Acids Res. 33*, 3659–3666.
55. Touboul, D. and Zenobi, R. (2009) A simple model for exploring conformation of highly-charged electrosprayed single-stranded oligonucleotides, *Chem. Comm. 2009*, 298–300.
56. Smith, S. I., Guziec, L. J., Guziec, F. S., Jr., Hasinoff, B. B., and Brodbelt, J. S. (2007) Evaluation of relative DNA binding affinities of anthrapyrazoles by electrospray ionization mass spectrometry, *J. Mass Spectrom. 42*, 681–688.
57. Kaltashov, I. A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal. Chem. 77*, 5370–5379.
58. Millard, J. T. (1999) Molecular probes of DNA structure, *Comp. Nat. Prod. Chem. 7*, 81–103.
59. Smith, S. I. and Brodbelt, J. S. (2010) Rapid characterization of cross-links, mono-adducts, and non-covalent binding of psoralens to deoxyoligonucleotides by LC-UV/ESI–MS and IRMPD mass spectrometry, *Analyst 135*, 943–952.
60. Parr, C., Pierce, S. E., Smith, S. I., and Brodbelt, J. S. (2011) Investigation of the reactivity of oligodeoxynucleotides with glyoxal and KMnO4 chemical probes by electrospray ionization mass spectrometry, *Int. J. Mass Spectrom. 304*, 115–123.
61. Hofstadler, S. A. and Griffey, R. H. (2001) Analysis of noncovalent complexes of DNA and RNA by mass spectrometry, *Chem. Rev. 101*, 377–390.
62. Hofstadler, S. A., Sannes-Lowery, K. A., and Griffey, R. H. (2000) Enhanced gas-phase hydrogen–deuterium exchange of oligonucleotide and protein ions stored in an external multipole ion reservoir, *J. Mass Spectrom. 35*, 62–70.
63. Mo, J., Todd, G. C., and Håkansson, K. (2009) Characterization of nucleic acid higher order structure by gas-phase H/D exchange in a quadrupole-FT–ICR mass spectrometer, *Biopolymers 91*, 256–264.
64. Baker, E. S., Dupuis, N. F., and Bowers, M. T. (2009) DNA hairpin, pseudoknot, and cruciform stability in a solvent-free environment, *J. Phys. Chem. B 113*, 1722–1727.
65. Rayburn, E. R. and Zhang, R. W. (2008) Antisense, RNAi, and gene silencing strategies for therapy: Mission possible or impossible?, *Drug Discov. Today 13*, 513–521.
66. Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, *Nature Chamun 391*, 806–811.
67. Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) Conformation of the Tar RNA-arginine complex by NMR-spectroscopy, *Science 257*, 76–80.
68. Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., and Williamson, J. R. (1992) Preparation of isotopically labeled

ribonucleotides for multidimensional NMR-spectroscopy of RNA, *Nucleic Acids Res. 20*, 4515–4523.

69. Fürtig, B., Richter, C., Wöhnert, J., and Schwalbe, H. (2003) NMR spectroscopy of RNA, *Chembiochem 4*, 936–962.
70. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 Ångstrom resolution, *Science 292*, 883–896.
71. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Ångstrom resolution, *Science 289*, 905–920.
72. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature (London) 407*, 327–339.
73. Gmeiner, W. H., Sahasrabudhe, P., and Pon, R. T. (1995) Use of shaped pulses for semi-selective excitation of imino H-1 resonances in duplex DNA and RNA, *Magnetic Reson. Chem. 33*, 449–452.
74. Nonin, S., Jiang, F., and Patel, D. J. (1997) Imino proton exchange and base-pair kinetics in the AMP-RNA aptamer complex, *J. Mol. Biol. 268*, 359–374.
75. Li, T. S., Johnson, J. E., and Thomas, G. J. (1993) Raman dynamic probe of hydrogen-exchange in bean pod mottle virus - base-specific retardation of exchange in packaged ssRNA, *Biophys. J. 65*, 1963–1972.
76. Ramstein, J. and Erdmann, V. A. (1981) A hydrogen-exchange study of escherichia-coli 5S RNA, *Nucleic Acids Res. 9*, 4081–4086.
77. Hofstadler, S. A., Sannes-Lowery, K. A., and Hannis, J. C. (2005) Analysis of nucleic acids by FTICR MS, *Mass Spectrom. Rev. 24*, 265–285.
78. Budnik, B. A., Haselmann, K. F., and Zubarev, R. A. (2001) Electron detachment dissociation of peptide di-anions: an electron-hole recombination phenomenon, *Chem. Phys. Lett. 342*, 299–302.
79. Yang, J., Mo, J. J., Adamson, J. T., and Hakansson, K. (2005) Characterization of oligodeoxynucleotides by electron detachment dissociation Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem. 77*, 1876–1882.
80. Dixon, W. J., Hayes, J. J., Levin, J. R., Weidner, M. F., Dombroski, B. A., and Tullius, T. D. (1991) Hydroxyl radical footprinting, *Methods Enzymol. 280*, 380–413.
81. Brenowitz, M., Chance, M. R., Dhavan, G., and Takamoto, K. (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical "footprinting", *Curr. Opin. Struct. Biol. 12*, 648–653.
82. Celander, D. W. and Cech, T. R. (1991) Visualizing the higher order folding of a catalytic RNA molecule, *Science 251*, 401–407.
83. Sclavi, B., Woodson, S., Sullivan, M., Chance, M. R., and Brenowitz, M. (1998) Time-resolved synchrotron X-ray 'footprinting', a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding, *J. Mol. Biol. 266*, 144–159.
84. Yu, E. and Fabris, D. (2003) Direct probing of RNA structures and RNA-protein interactions in the HIV-1 packaging signal by chemical modification and electrospray ionization fourier transform mass spectrometry, *J. Mol. Biol. 330*, 211–223.
85. Kellersberger, K. A., Yu, E., Kruppa, G. H., Young, M. M., and Fabris, D. (2004) Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation, *Anal. Chem. 76*, 2438–2445.
86. Turner, K. B., Yi-Brunozzi, H. Y., Brinson, R. G., Marino, J. P., Fabris, D., and Le Grice, S. F. J. (2009) SHAMS: Combining chemical modification of RNA with mass spectrometry to examine polypurine tract-containing RNA/DNA hybrids, *RNA 15*, 1605–1613.
87. Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry, *Proc. Natl. Acad. Sci. USA 97*, 5802–5806.
88. Fabris, D. and Yu, E. T. (2010) Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D, *J. Mass Spectrom. 45*, 841–860.
89. Fabris, D. (2010) A role for the MS analysis of nucleic acids in the post-genomics age, *J. Am. Soc. Mass Spectrom. 21*, 1–13.
90. Yu, E. T., Hawkins, A., Eaton, J., and Fabris, D. (2008) MS3D structural elucidation of the HIV-1 packaging signal, *Proc. Natl. Acad. Sci. USA 105*, 12248–12253.
91. Turner, K. B., Kohlway, A. S., Hagan, N. A., and Fabris, D. (2009) Noncovalent probes for the investigation of structure and dynamics of protein–nucleic acid assemblies: the case of NC-mediated dimerization of genomic RNA in HIV-1, *Biopolymers 91*, 283–296.
92. Turner, K. B., Brinson, R. G., Yi-Brunozzi, H. Y., Rausch, J. W., Miller, J. T., Le Grice, S. F. J., Marino, J. P., and Fabris, D. (2008) Structural probing of the HIV-1 polypurine tract RNA: DNA hybrid using classic nucleic acid ligands, *Nucleic Acids Res. 36*, 2799–2810.
93. Fabris, D., Kellersberger, K. A., and Wilhilde, J. A. (2011) Higher order structure of nucleic acids in the gas phase: top-down analysis of base-pairing interactions, *Int. J. Mass Spectrom.*, In press.
94. Deterding, L. J., Kast, J., Przybylski, M., and Tomer, K. B. (2000) Molecular characterization of a tetramolecular complex between dsDNA and a DNA-binding leucine zipper peptide dimer by mass spectrometry, *Bioconjug. Chem. 11*, 335–344.
95. Akashi, S., Osawa, R., and Nishimura, Y. (2005) Evaluation of protein-DNA binding affinity by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 16*, 116–125.
96. Bayer, E., Bauer, T., Schmeer, K., Bleicher, K., Maler, M., and Gaus, H. J. (1994) Analysis of double-stranded oligonucleotides by electrospray mass-spectrometry, *Anal. Chem. 66*, 3858–3863.
97. Hoyne, P. R., Benson, L. M., Veenstra, T. D., Maher, L. J., and Naylor, S. (2001) RNA–RNA noncovalent interactions investigated by microspray ionization mass spectrometry, *Rapid Commun. Mass Spectrom. 15*, 1539–1547.

98. Turner, K. B., Hagan, N. A., and Fabris, D. (2007) Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein, *J. Mol. Biol. 369*, 812–828.
99. Zhou, M., Sandercock, A. M., Fraser, C. S., Ridlova, G., Stephens, E., Schenauer, M. R., Yokoi-Fong, T., Barsky, D., Leary, J. A., Hershey, J. W., Doudna, J. A., and Robinson, C. V. (2008) Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3, *Proc. Natl. Acad. Sci. USA 105*, 18139–18144.
100. Fabris, D. (2000) Steady-state kinetics of ricin A-chain reaction with the sarcin-ricin loop and with HIV-1 Psi-RNA hairpins evaluated by direct infusion electrospray ionization mass spectrometry, *J. Am. Chem. Soc. 122*, 8779–8780.
101. Hofstadler, S. A., Sannes-Lowery, K. A., Crooke, S. T., Ecker, D. J., Sasmor, H., Manalili, S., and Griffey, R. H. (1999) Multiplexed screening of neutral mass-tagged RNA targets against ligand libraries with electrospray ionization FTICR MS: A paradigm for high-throughout affinity screening, *Anal. Chem. 71*, 3436–3440.
102. Mei, H. Y., Mack, D. P., Galan, A. A., Halim, N. S., Heldsinger, A., Loo, J. A., Moreland, D. W., SannesLowery, K. A., Sharmeen, L., Truong, H. N., and Czarnik, A. W. (1997) Discovery of selective, small-molecule inhibitors of RNA complexes.1. The Tat protein TAR RNA complexes required for HIV-1 transcription, *Bioorg. Med. Chem. 5*, 1173–1184.
103. Seth, P. P., Miyaji, A., Jefferson, E. A., Sannes-Lowery, K. A., Osgood, S. A., Propp, S. S., Ranken, R., Massire, C., Sampath, R., Ecker, D. J., Swayze, E. E., and Griffey, R. H. (2005) SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain, *J. Med. Chem. 48*, 7099–7102.
104. Sannes-Lowery, K. A., Griffey, R. H., and Hofstadler, S. A. (2000) Measuring dissociation constants of RNA and aminoglycoside antibiotics by electrospray ionization mass spectrometry, *Anal. Biochem. 280*, 264–271.
105. Hagan, N. and Fabris, D. (2003) Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 psi-recognition element, *Biochemistry 42*, 10736–10745.
106. Costello, C. E., Perrault, H., and Ngoka, L. C. (1996) Mass spectrometric and tandem mass spectrometric approaches to the analysis of carbohydrates, In *Mass spectrometry in the biological sciences* (Burlingame, A. L.,and Carr, S. A., Eds.), pp 365–384, Humana Press, Totowa, N.J.
107. Zaia, J. (2004) Mass spectrometry of oligosaccharides, *Mass Spectrom. Rev. 23*, 161–227.
108. Biemann, K., and Martin, S. A. (1987) Mass spectrometric determination of the amino acid sequence of peptides and proteins, *Mass Spectrom. Rev. 6*, 1–75.
109. Domon, B. and Costello, C. E. (1988) A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates, *Glycoconj. J. 5*, 397–409.
110. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a fourier transform mass spectrometer, *Anal. Chem. 71*, 4431–4436.
111. Marino, K., Bones, J., Kattla, J. J., and Rudd, P. M. (2010) A systematic approach to protein glycosylation analysis: a path through the maze, *Nat. Chem. Biol. 6*, 713–723.
112. Zaia, J. (2004) Mass spectrometry of oligosaccharides, *Mass Spectrom. Rev. 23*, 161–227.
113. Zaia, J. (2010) Mass spectrometry and glycomics, *OMICS 14*, 401–418.
114. Dreyfuss, J., Jacobs, C., Gindin, Y., Benson, G., Staples, G., and Zaia, J. (2011) Targeted analysis of glycomics liquid chromatography/mass spectrometry data, *Anal. Bioanal. Chem. 399*, 727–735.
115. Sletmoen, M. and Stokke, B. T. (2008) Review: Higher order structure of (1,3)-beta-D-glucans and its influence on their biological activities and complexation abilities, *Biopolymers 89*, 310–321.
116. Rudd, T. R., Skidmore, M. A., Guerrini, M., Hricovini, M., Powell, A. K., Siligardi, G., and Yates, E. A. (2010) The conformation and structure of GAGs: recent progress and perspectives, *Curr. Opin. Struct. Biol. 20*, 567–574.
117. Guttman, M., Scian, M., and Lee, K. K. (2011) Tracking hydrogen/deuterium exchange at glycan sites in glycoproteins by mass spectrometry, *Anal. Chem.*, *83*, 7492–7499.
118. Champetier, G. and Viallard, R. (1938) Reaction d'echange de la cellulose et de l'eau lourde. Hydratation de la cellulose, *B. Soc. Chim. 33*, 1042–1048.
119. Price, N. P. J. (2006) Oligosaccharide structures studied by hydrogen-deuterium exchange and MALDI–TOF mass spectrometry, *Anal. Chem. 78*, 5302–5308.
120. Seo, Y., Schenauer, M. R., and Leary, J. A. (2011) Biologically relevant metal-cation binding induces conformational changes in heparin oligosaccharides as measured by ion mobility mass spectrometry, *Int. J. Mass Spectrom. 303*, 191–198.
121. Abzalimov, R. R., Dubin, P. L., and Kaltashov, I. A. (2007) Glycosaminoglycans as naturally occurring combinatorial libraries: Developing a mass spectrometry-based strategy for characterization of anti-thrombin interaction with low molecular weight heparin and heparin oligomers, *Anal. Chem. 79*, 6055–6063.
122. Hyland, R. M., Sun, J., Griener, T. P., Mulvey, G. L., Klassen, J. S., Donnenberg, M. S., and Armstrong, G. D. (2008) The bundlin pilin protein of enteropathogenic Escherichia coli is an *N*-acetyllactosamine-specific lectin, *Cel & Microbiol. 10*, 177–187.
123. El-Hawiet, A., Kitova, E. N., Kitov, P., Eugenio, L., Ng, K. K. S., Mulvey, G. L., Dingle, T. C., Szpacenko, A., Armstrong, G. D., and Klassen, J. S. (2011) Binding of *Clostridium difficile* toxins to human milk oligosaccharides, *Glycobiology 21*, 1217–1227.
124. Zhou, C. Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z. G., and Janin, J. (2001) Silk fibroin: structural implications of a remarkable amino acid sequence, *Proteins 44*, 119–122.
125. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., and Savage, K. N. (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function, *J. Exp. Biol. 202*, 3295–3303.
126. Valluzzi, R., Winkler, S., Wilson, D., and Kaplan, D. L. (2002) Silk: molecular organization and control of

assembly, *Philos. Trans. R. Soc. Lond B Biol. Sci. 357*, 165–167.

127. Gelse, K., Poschl, E., and Aigner, T. (2003) Collagens - structure, function, and biosynthesis, *Adv. Drug Del. Rev. 55*, 1531–1546.
128. Ottani, V., Martini, D., Franchi, M., Ruggeri, A., and Raspanti, M. (2002) Hierarchical structures in fibrillar collagens, *Micron 33*, 587–596.
129. O'Sullivan, A. C. (1997) Cellulose: the structure slowly unravels, *Cellulose 4*, 173–207.
130. Kadla, J. F. and Gilbert, R. D. (2000) Cellulose structure: A review, *Cell. Chem. Technol. 34*, 197–216.
131. Staudinger, H. (1932) *Die Hochmolekularen organischen Verbindungen, Kautschuk und Cellulose*, J. Springer, Berlin.
132. Staudinger, H. (1947) *Makromolekulare Chemie und Biologie*, Wepf, Basel.
133. Carothers, W. H., Mark, H. F., and Whitby, G. S. (1940) *Collected papers by Wallace Hume Carothers on high polymeric substances*, Interscience publishers inc., New York.
134. Barron, A. E. and Zuckerman, R. N. (1999) Bioinspired polymeric materials: in-between proteins and plastics, *Curr. Opin. Chem. Biol. 3*, 681–687.
135. Cunliffe, D., Pennadam, S., and Alexander, C. (2004) Synthetic and biological polymers - merging the interface, *Eur. Polym. J. 40* 5–25.
136. van Hest, J. C. M. and Tirrell, D. A. (2001) Protein-based materials, toward a new level of structural control, *Chem. Commun.2001* 1897–1904.
137. O'Brien, J. P., Fahnestock, S. R., Termonia, Y., and Gardner, K. C. H. (1998) Nylons from nature: Synthetic analogs to spider silk, *Adv. Mater. 10*, 1185–1195.
138. Hinman, M. B., Jones, J. A., and Lewis, R. V. (2000) Synthetic spider silk: a modular fiber, *Trends Biotechnol. 18*, 374–379.
139. Bini, E., Knight, D. P., and Kaplan, D. L. (2004) Mapping domain structures in silks from insects and spiders related to protein assembly, *J. Mol. Biol. 335*, 27–40.
140. Roberts, M. J., Bentley, M. D., and Harris, J. M. (2002) Chemistry for peptide and protein PEGylation, *Adv. Drug Del. Rev. 54*, 459–476.
141. Greenwald, R. B., Choe, Y. H., McGuire, J., and Conover, C. D. (2003) Effective drug delivery by PEGylated drug conjugates, *Adv. Drug Del. Rev. 55*, 217–250.
142. Matthews, C. N. (1994) Hardware and software in biology - simultaneous origin of proteins and nucleic-acids via hydrogen-cyanide polymers, *J. Biol. Phys. 20*, 275–281.
143. Abzalimov, R. R., Frimpong, A., and Kaltashov, I. A. (2011) Structural characterization of protein-polymer conjugates. I. Assessing heterogeneity of a small PEGylated protein and mapping conjugation sites using ion exchange chromatography and top-down tandem mass spectrometry, *Int. J. Mass Spectrom. (in press).*
144. Monkarsh, S. P., Ma, Y., Aglione, A., Bailon, P., Ciolek, D., Debarbieri, B., Graves, M. C., Hollfelder, K., Michel, H., Palleroni, A., Porter, J. E., Russoman, E., Roy, S., and Pan, Y.-C. E. (1997) Positional isomers of monopegylated interferon α-2a: Isolation, characterization, and biological activity, *Anal. Biochem. 247*, 434–440.
145. Foser, S., Schacher, A., Weyer, K. A., Brugger, D., Dietel, E., Marti, S., and Schreitmüller, T. (2003) Isolation, structural characterization, and antiviral activity of positional isomers of monopegylated interferon α-2a (PEGASYS), *Prot. Exp. Purif. 30*, 78–87.
146. Arduini, R. M., Li, Z., Rapoza, A., Gronke, R., Hess, D. M., Wen, D., Miatkowski, K., Coots, C., Kaffashan, A., Viseux, N., Delaney, J., Domon, B., Young, C. N., Boynton, R., Chen, L. L., Chen, L., Betzenhauser, M., Miller, S., Gill, A., Pepinsky, R. B., Hochman, P. S., and Baker, D. P. (2004) Expression, purification, and characterization of rat interferon-β, and preparation of an N-terminally PEGylated form with improved pharmacokinetic parameters, *Prot. Exp. Purif. 34*, 229–242.
147. Yoo, C., Suckau, D., Sauerland, V., Ronk, M., and Ma, M. H. (2009) Toward top-down determination of PEGylation site using MALDI in-source decay MS analysis, *J. Am. Soc. Mass Spectrom. 20*, 326–333.
148. Lu, X. J., Gough, P. C., DeFelippis, M. R., and Huang, L. H. (2010) Elucidation of PEGylation site with a combined approach of in-source fragmentation and CID MS/MS, *J. Am. Soc. Mass Spectrom. 21*, 810–818.
149. Houde, D., Berkowitz, S. A., and Engen, J. R. (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies, *J. Pharm. Sci. 100*, 2071–2086.
150. Nakano, T. and Okamoto, Y. (2001) Synthetic helical polymers: Conformation and function, *Chem. Rev. 101*, 4013–4038.
151. Zagar, E. and Grdadolnik, J. (2003) An infrared spectroscopic study of H-bond network in hyperbranched polyester polyol, *J. Mol. Struct. 658*, 143–152.
152. Bosman, A. W., Janssen, H. M., and Meijer, E. W. (1999) About dendrimers: Structure, physical properties, and applications, *Chem. Rev. 99*, 1665–1688.
153. Mong, T. K. K., Niu, A. Z., Chow, H. F., Wu, C., Li, L., and Chen, R. (2001) beta-alanine-based dendritic beta-peptides: Dendrimers possessing unusually strong binding ability towards protic solvents and their self-assembly into nanoscale aggregates through hydrogen-bond interactions, *Chem. Eur. J. 7*, 686–699.
154. Adhiya, A. and Wesdemiotis, C. (2002) Poly(propylene imine) dendrimer conformations in the gas phase: a tandem mass spectrometry study, *Int. J. Mass Spectrom. 214*, 75–88.
155. Eichman, J. D., Bielinska, A. U., Kukowska-Latallo, J. F., and BakerJr, J. R. (2000) The use of PAMAM dendrimers in the efficient transfer of genetic material into cells, *Pharm. Sci. Technol. Today 3*, 232–245.

9

MASS SPECTROMETRY ON THE FRONTIERS OF MOLECULAR BIOPHYSICS AND STRUCTURAL BIOLOGY: PERSPECTIVES AND CHALLENGES

It is now indisputable that mass spectrometry (MS) has become a routine analytical tool in diverse areas of molecular biophysics and structural biology. Numerous examples presented in the previous chapters attest to the highly visible (and widely appreciated) role currently played by MS. As appreciation for the technique grows, so do the expectations. Increasingly, MS is being looked at with the hope of providing long-sought analytical solutions to deal with complex and challenging problems, ranging from behavior of highly heterogeneous systems to protein aggregation, to name a few. This chapter focuses on several areas where MS is either currently making progress or shows tremendous promise. We will open the list with a discussion of MS based strategies that target structure and dynamics of a notoriously difficult class of biopolymers: membrane proteins. Then we discuss protein aggregation, which has been a hot topic in biophysical research in recent years, not only due to its intimate involvement in etiology of "conformational diseases" (e.g., Alzheimer's or Parkinson's), but also due to its importance in the burgeoning biotechnology and biopharmaceutical sectors. Next we briefly consider a relatively new theme in MS research: The problem of structural heterogeneity, which has also come to prominence due to its extreme importance in the field of protein therapeutics. We close this chapter with a discussion of new experimental approaches whose design was influenced by the increasing trend to break away from the reductionist paradigm and embrace the complexity of living systems.

9.1. MASS SPECTROMETRY AND THE UNIQUE CHALLENGES OF MEMBRANE PROTEINS

Membrane proteins constitute a broad class of biopolymers, whose amphiphilic nature makes them notoriously difficult for structural and functional analysis at the whole protein level (1–3). By some counts, membrane proteins constitute one-third of all proteins, but only a tiny fraction of them have been characterized well enough to produce high-resolution structure (as of July 2011, structures were known for only 297 unique proteins according to the MPSTRUC database[*]). In most cases, membranes are essential for the integrity of the proteins they host, although the structure and dynamic properties of membrane proteins are actually defined by a wide spectrum of intermolecular forces. These include protein interactions with the hydrophobic "bulk" of the membrane, its polar interface region, internal and external water molecules, as well as interactions between various segments of the protein itself (Fig. 9.1). Such amphipathic character of membrane proteins results in their general insolubility, which makes any experimental studies extremely difficult. Mass spectrometry is not an exception, since even sequencing of membrane proteins is often problematic due to their extreme instability in solutions that are commonly used in MS work (6).

[*] http://blanco.biomol.uci.edu/mpstruc/listAll/list

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

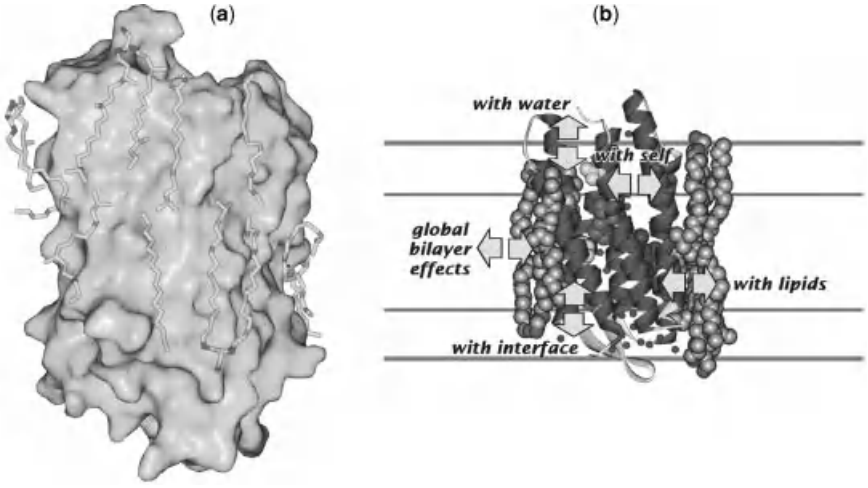


Figure 9.1. (*a*) A view of membrane protein interactions with lipids. Native lipids are seen bound to the surface of bacteriorhodopsin and suggest intimate and specific interactions between the protein and lipids. [Reproduced with permission from (4)] (*b*) A schematic representation of the polypeptide interactions that determine the structure and stability of membrane proteins. The horizontal black lines represent schematically the total thickness of the lipid bilayer, and the horizontal gray lines the central hydrocarbon core bounded by the interfacial regions. Besides interactions of the polypeptide chain with itself, water, neighboring lipids, and the membrane interface, the thermodynamic and electrostatic properties of the lipid bilayer itself are important. The lipid bilayer, like proteins, resides in a free energy minimum resulting from numerous interactions. This equilibrium can be disturbed by the introduction of proteins or other solutes, resulting in the so-called bilayer effects, which also include solvent properties peculiar to bilayers that arise from motional anisotropy and chemical heterogeneity. [Reproduced with permission from (5). Copyright © 2001 American Society for Biochemistry and Molecular Biology.]

9.1.1. Analysis of Membrane Proteins in Organic Solvents

Earlier attempts at addressing the solubility problem used organic solvents to mimic the nonpolar environment of the membrane interior. Although the simplicity of this approach is very appealing, it is probably unrealistic to expect that removing it from the membrane and placing it in a homogeneous nonpolar environment would not compromise protein behavior. Although the secondary structure may often be retained, a significant, proportion of the protein tertiary structure will likely be perturbed or even completely lost (7). Nevertheless, this approach has provided some useful information on the structure and behavior of several membrane proteins. For example, Dobson and co-workers (8) used a small transmembrane channel peptide gramicidin A to explore influence of the environment on its conformation. The "channel" conformation of the peptide adopted in lipid membranes and sodium dodecyl sulfate (SDS) micelles is a right-handed $\beta^{6.3}$ helix dimer, although other conformations have been reported as well. The exact conformation assumed by the peptide in lipid bilayers remains a subject of debate (9,10). Furthermore, it is still unclear what conformation is adopted by gramicidin A in polar solvents and whether or not it retains its dimer-forming ability in such solutions. Mass spectral data provided strong evidence that gramicidin A remained a monomer in 2,2,2-trifluoroethanol (TFE) and dimethyl sulfoxide (DMSO) solutions, while the existence of the dimeric species became evident in ethanol solutions (8). Despite its inability to form dimer in TFE, the peptide was apparently highly structured in this environment, as suggested by significant backbone amide protection revealed by hydrogen–deuterium exchange (HDX) MS measurements (8). These measurements were rather remarkable, as exchange was carried out in anhydrous solvent. The HDX reactions in TFE were initiated by diluting a concentrated TFE (CF_3 CH_2OH) solution of gramicidin A in d_1-TFE (CF_3CH_2OD) 1:99 (*v*:*v*). Exchange of a reference tripeptide (Ala_3) was also monitored under these conditions to provide a measure of the intrinsic exchange rate in TFE.

In another study, Chitta and Gross (11) used electrospray ionization (ESI) MS to characterize gramicidin A dimerization in solutions ranging from relatively hydrophilic (TFE) to very hydrophobic (*n*-propanol). The degree of dimerization was clearly correlated with the dielectric constant of the solvent. Careful evaluation of the ESI MS data led the authors to a conclusion that up to 70% of the noncovalently bound dimers formed in solution could be preserved in the gas phase under the most favorable conditions. Fluorinated alcohols (TFE and hexafluoroisopropanol, HFIP) were also used by Waring and co-workers (12) to examine structure of a viral fusion peptide (*N*-terminal

peptide FP of human immunodeficiency virus HIV-1 gp41) under conditions approximating membrane-like environments. The amide protection pattern was evaluated in a site-specific fashion using HDX in solution (50% TFE or 70% HFIP) and peptide ion fragmentation in the gas phase. The resulting map of fusion peptide secondary structure compared favorably with the earlier nuclear magnetic resonance (NMR) studies carried out in similar environments mimicking membranes (12).

In some cases, stabilization of membrane proteins can be achieved by covalent modification. For example, Poschner and Langosch, used covalent attachment of lipids to prevent protein misfolding and aggregation (13). This method, however, cannot be adopted universally, and typically two approaches are commonly used to deal with the solubility problem. One relies on using detergents to isolate, solubilize, and manipulate membrane proteins (4), while the other utilizes lipid vesicles or other membrane mimics as membrane surrogates (14,15). Both approaches, as well as select examples of relevant studies that use MS as a biophysical tool, are briefly considered in the following sections.

9.1.2. Analysis of Membrane Proteins Using Detergents

Detergents are amphipathic molecules that act as surfactants, and are indispensable as solubilizing agents for membrane proteins (16). Most detergents belong to one of the following groups: ionic, nonionic, or zwitterionic. Several popular detergents are listed in Table 9.1. Detergents have a natural tendency to self-associate and to interact with other molecules, and exhibit a wide range of behavior and assembled structures depending on their concentration, solution conditions (pH, ionic strength, etc.), as well as the presence of proteins and other molecules (Fig. 9.2). Self-association (leading to micelle formation) occurs at a threshold concentration, usually termed *critical micelle concentration* (CMC). Micelle size is usually characterized by an aggregation number (the average number of detergent monomers per micelle). The shape of the micelle is determined by the packing of the monomer tails and is dependent on the aggregation number. It is usually quite nonuniform, despite the commonly used notion of an ideal spherical micelle (4). Packing defects result in considerable contact between hydrophobic tails and water. Micelles are very dynamic structures and rapidly exchange micellar components with the solvent. A detergent is capable of solubilizing amphipathic molecules only at a concentration that exceeds the CMC. Solubilization of many integral membrane proteins actually occurs significantly above the CMC level, since the detergent molecules also interact with the hydrophobic surfaces of the membrane protein to create protein-detergent complexes (4).

A major technical problem that is often encountered when MS is used to characterize detergent-solubilized membrane proteins is the suppressive effect of detergents (18). Sometimes it is possible to substitute detergents with organic solvent, in which the protein would not precipitate (19–21). Typically, this is accomplished by protein precipitation, followed by removal of the detergent, and resolubilization of the protein in a suitable nonpolar solvent. Separation of membrane proteins from detergents can also be accomplished in a single step using high-performance liquid chromatography (HPLC), which is compatible with direct mass spectrometric analysis [e.g., online HPLC–ESI MS (22)]. Reverse-phase chromatography is normally expected to denature proteins, however, there is substantial evidence that this is not always the case (22). Another method to obtain high-quality mass spectra of detergent-solubilized membrane proteins avoids the protein precipitation and resolubilization steps by utilizing direct protein extraction into a nonpolar solvent (23). Finally, a group of ESI friendly surfactants [e.g., perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOSA)] have

TABLE 9.1. Properties of Commonly Used Detergents.[a]

Name	Monomeric Mass	Critical Micelle Concentration	Aggregation Number	Comments
SDS	288	1.2–7.1 mM	62–101	Ionic detergent, strongly denaturing
3-[(3-Cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS)	615	3–10 mM	4–14	Steroid-based detergent, relatively mild
Dodecyldimethyl-*N*-amineoxide (DDAO)	229	2.2 mM	69–73	Zwitterionic detergent (uncharged at pH > 5)
N-Dodecylphosphocholine (DPC)	352	1.1 mM	50–60	Ionic detergent, an efficient solubilizer of hydrophobic or amphipathic α-helices. Dynamic kehavior of DPC at low temperatures corresponds to that in a phosphotidylcholine membranewater interface above its room temperature
*p-tert*C8ØE$_{<9.5>}$ (Triton X-100)	625	0.25 mM	75–165	Polyoxyethylene glycol detergent, generally mild
C_{12} sorbitan E$_{<9.5>}$ (Tween-20)	1240	60 μM		

[a] Based on data compiled in (16)

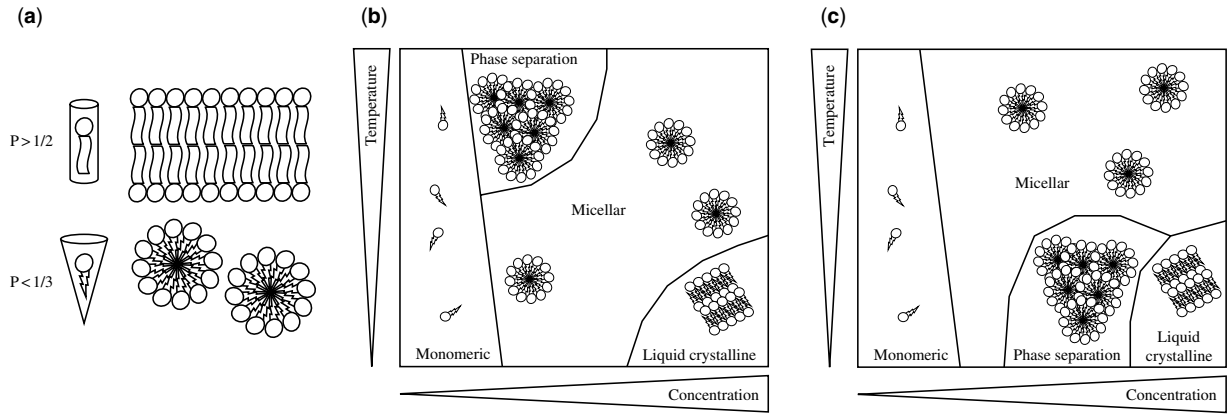


Figure 9.2. (*a*) Self-association of detergents in aqueous solutions can lead to formation of bilayers or micelles depending on the shape of the detergent molecule. It is best described by the packing parameter calculated as $P = (\text{detergent chain volume})/(\text{headgroup cross-section} \times \text{hydrophobic chain length})$. (*b*) Simplified phase diagram for a detergent with a lower consolute boundary (most nonionic detergents fall into this group). (***c***) Simplified phase diagram for a detergent with an upper consolute boundary (a number of glycosidic and zwitterionic detergents fall into this group). Note that the liquid-crystalline phase can also consist of hexagonally packed cylindrical micelles. [Adapted from (17)].

been reported (24), but their current use is limited mostly to online separation (LC or electrophoresis) with MS detection, where they are employed as ion-pairing agents. In addition, there are several methods of detergent removal particularly suitable for matrix-assisted laser desorption ionization (MALDI) MS that make use of phase separation during the crystallization step (25). It is commonly accepted that only relatively mild (nonionic or zwitterionic) detergents should be used in the analyses of membrane proteins by MS. However, there are reports of MALDI MS analyses of protein samples containing strong (ionic) detergents, (e.g. SDS (26), as well as its milder analogue, ammonium dodecyl sulfate (ADS) (27).

As already discussed, even though a large number of membrane proteins can be solubilized in ESI friendly nonpolar solvents, their structures are unlikely to be preserved in such environments. This dilemma (inability of the electrospray-friendly solvent systems to maintain the integrity of membrane proteins and inability of ESI to handle membrane-friendly solvent systems) has presented a seemingly insurmountable problem for analysis of membrane proteins by ESI MS. An elegant solution to this problem was offered by Griffiths and co-workers (28), who were able to carry out direct ESI MS analysis of a noncovalently bound membrane protein complex (trimer) microsomal glutathione transferase-1 in its native state after reconstituting it in a minimum amount of detergent. The detergent used in this study was Triton X-100, which was apparently well tolerated by ESI MS, although the increase of the detergent content of the sample inevitably led to deterioration of spectral quality; tolerable levels of Triton X-100 were found to be $<0.1\%$. A range of other detergents was tested by the authors of this study (MEGA-10, Thesit and Zwittergent 3–12), however, they were found to be less compatible with ESI MS compared with Triton X-100. To avoid excessive protein–detergent adduct formation, relatively high collisional activation of ions in the ESI interface region was used to induce dissociation of adduct ions and increase signal-to-noise ratio (S/N) (28). Some trimeric ions survived such harsh conditions in the ESI interface, although the mass spectrum was clearly dominated by the signal corresponding to the monomer ions, emphasizing the marginal stability of the trimeric protein complex in the gas phase. Nevertheless, the gas-phase origin of the monomer ions was rather obvious due to the anomalous extent of multiple charging caused by asymmetric charge partitioning upon complex dissociation, a phenomenon discussed in Section 4.1.3.

A similar approach was later applied by Robinson and co-workers (29) to study the structure of large transmembrane proteins. A heterotetrameric protein adenosine triphosphate (ATP) binding cassette transporter ($BtuC_2D_2$) was stabilized in solution within *n*-dodecyl-β-D-maltoside (DDM) micelles followed by nano-ESI MS analysis. The entire protein complex was maintained largely intact in the gas phase when mild desolvation conditions were used in the ESI interface. Unfortunately, insufficient desolvation resulted in formation of multiple (>100) detergent adducts, which resulted in poor quality MS data and prevented individual complex ion species from being resolved and identified (Fig. 9.3*a*). However, increasing the efficiency of protein desolvation in the ESI interface resulted in a dramatic improvement of spectral quality by removing DDM adducts from the protein complex ions, and allowed their masses to be determined with good precision (Fig. 9.3*b*). Further increase of the efficiency of collisional activation in the ESI interface led to complex dissociation; once again,

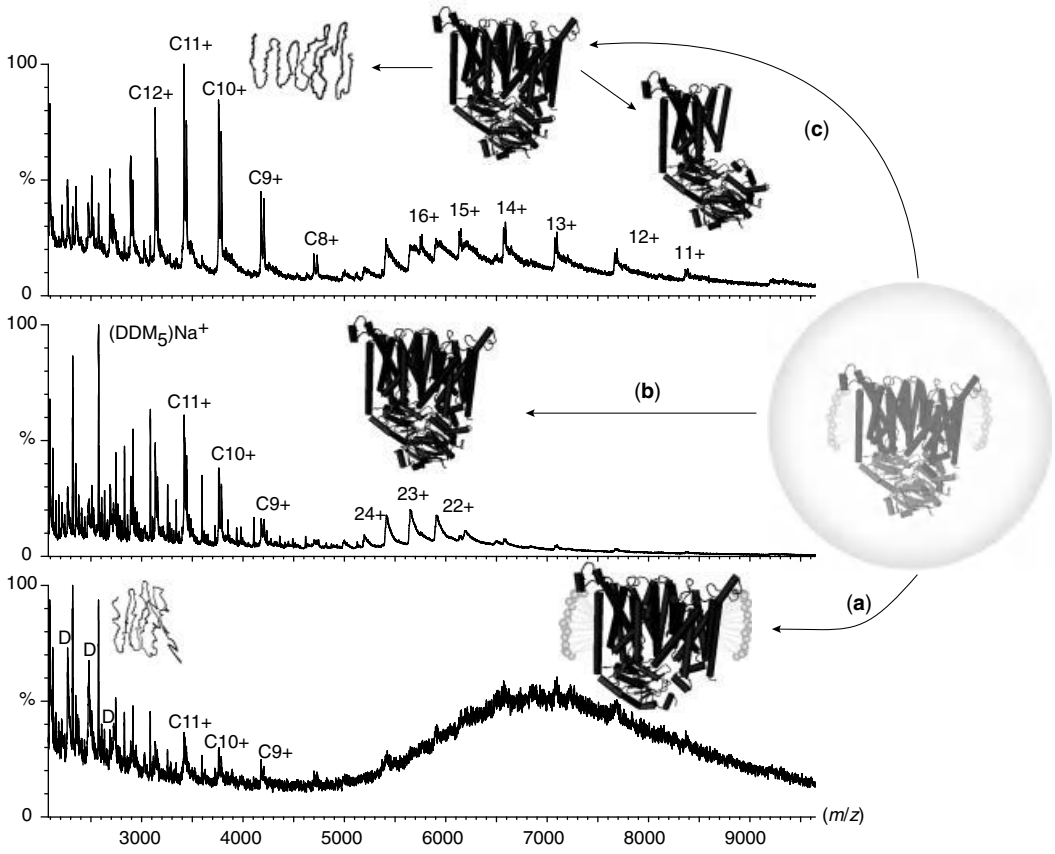


Figure 9.3. Schematic representation of the emergence of the intact membrane complex $BtuC_2D_2$ from a micelle contained within an electrospray-generated droplet and subsequent gas-phase dissociation pathways. Populations of ions corresponding to the protein complex associated with aggregates of detergent molecules are observed above *m*/*z* 5000. At low *m*/*z*, the dominant dissociation product is the unfolded BtuD subunit [pathway (*a*)]. Increasing the number of collisions leads to the release of the intact tetramer [pathway (*b*)]. Further increases in the number of collisions lead to the dissociation of BtuC and formation of a trimer [pathway (*c*)]. [Adapted from (29).] (See color version of the figure in Color Plate section)

the occurrence of this process was clearly evident due to the extreme charging patterns exhibited by the dissociation products (Fig. 9.3*c*). Another interesting feature of this work is that the authors noticed that the extent of multiple charging observed in ESI mass spectra fell off the charge–surface correlation plot constructed for water-soluble proteins (30), which we already discussed (Fig. 4.7). This was suggested to be the result of partial shielding of the protein surface by detergent molecules, which prevented these regions from accommodating charges upon transition to the gas phase. Once the surface of membrane-embedded area was subtracted from the total protein surface, a much better agreement with the original charge–surface correlation plot was achieved (29). A continuation of this study employed IM–MS as a means of obtaining additional information on the protein geometry in solution, although the collisional cross-section of the $BtuC_2D_2$ did not have a clear correlation with that calculated based on the native structure of this protein, highlighting the possibility of both inter- and intrasubunit rearrangements in the gas phase (31).

Stoichiometry of detergent-stabilized membrane proteins has also been studied by a recently introduced modification of MALDI, called laser-induced liquid-beam (bead) desorption and ionization, LILBID (32). In this approach, the liquid matrix consists of a 10 μm diameter free liquid filament in vacuum (or a free droplet) that is excited with an infrared (IR) laser pulse (2.8 micron, which matches the absorption band of the O–H stretch vibration of bulk water). This method produces low charge-density macromolecular ions, although the mechanism remains poorly understood. This new technique had been shown to preserve noncovalent protein complexes, and also to be surprisingly tolerant to the presence of salts, buffers, and detergents (32). This makes it an attractive alternative to ESI MS for the analysis of detergent-stabilized membrane proteins. In fact, it has been successfully used to analyze the subunit stoichiometry of a purified tridecameric *c* ring of bacterial ATP synthase (33).

More recently, this new technique was applied to study the proton-pump bacteriorhodopsin and a potassium channel, KcsA (34); in both cases the stoichiometry found by LILBID reflects the known crystal structures. Intriguingly, at least for some membrane proteins, LILBID demonstrated a dependence of structure on the choice of detergent (34), a point that we will discuss further at the end of this section.

While the direct MS measurements discussed above provide information on the composition and stoichiometry of membrane protein complexes, other structural aspects of detergent-stabilized proteins can be probed by chemical labeling and HDX MS. In addition to being a great stumbling block in native ESI MS measurements, the challenges of handling the detergent in MS work flow have also made it difficult to deploy the highly successful HDX MS methods toward elucidating the structure of the detergent-stabilized membrane proteins. Relatively mild detergents, that are commonly used to keep the membrane protein in solution, cannot be easily precipitated, and even though they can be separated from the proteolytic peptides during the LC separation step, the residual detergent peaks frequently overlap with the peptide peaks and complicate data analysis in HDX MS experiments. To circumvent this problem, Forest and co-workers (35) have recently developed a method based on solid-phase extraction with chlorinated solvents that fully removes the detergent, which allowed HDX MS to be applied to study conformational dynamics of an integral membrane protein mitochondrial ADP/ATP carrier isoform 1.

Selective chemical modification (a technique discussed in detail in Section 4.3) is perhaps the most common MS based method of probing higher order structure of detergent-stabilized membrane proteins. An interesting example of such studies was presented by Whitelegge et al. (36,37) in a series of reports aimed at elucidating the details of protein–substrate interactions for a 47 kDa transmembrane protein, lactose permease, from *Escherichia coli*. Initially, protein modification by *N*-ethylmaleimide (targeting thiol groups), followed by ESI MS detection and identification of the reaction products, was used to probe changes of the microenvironments of cysteine residues induced by substrate binding (36). Treatment of the native protein solubilized in detergent micelles reveals only two reactive thiol groups out of eight Cys residues. Both Cys residues become protected in the presence of D-galactopyranosyl β-D-thiogalactopyranoside (TDG), a substrate analogue. Interestingly, labeling of one of these two Cys residues, which is a component of the substrate binding site, according to the model of the protein–substrate interaction (see Fig. 9.4) is inhibited completely in the presence of the substrate analogue. Significantly reduced (but not completely eliminated) alkylation of the second reactive Cys residue in the presence of the substrate analogue reflects a long-range conformational change caused by binding of the substrate (36). The molecular model of lactose permease (38) also predicts that the substrate interacts with Glu^{269} via a hydroxyl group in the galactopyranosyl ring (Fig. 9.4c). Covalent modification of carboxyl groups with carbodiimide followed by identification of the reaction products with ESI MS provides strong evidence that the substrate protects the protein against carbodiimide reactivity (37). A significant proportion of the decrease in reactivity occurs specifically in a nonapeptide containing Glu^{269}, while the reactivity of a mutant (E269D, which exhibits significantly lower affinity toward the substrate) is unaffected by the substrate (37). Monitoring the ability of different substrate analogues to protect against carbodiimide modification of Glu^{269} provided evidence that the C3 hydroxyl group of the galactopyranosyl ring (Fig. 9.4*c*) plays an important role in specificity, possibly by hydrogen-bonding with Glu^{269} (37).

Nonspecific oxidative labeling is another approach that has been used successfully in recent years as a means of probing conformation and dynamics of detergent-stabilized membrane proteins. For example, Chance and co-workers (39) used radiolytic footprinting (see Section 4.3 for a discussion of this technique) to probe the structure of two distinct conformations of a potassium channel; the differences observed between these two states were used to identify local conformational changes that occur during channel gating.

A very creative approach to probing contact topology of assemblies of membrane proteins was presented by Przybylski and co-workers (40), who used proteolysis in the presence of a strong detergent SDS, to identify protein–protein contact regions. A complex formed by an ion-channel protein (mitochondrial porin) and its ligand (adenine nucleotide translocator) was separated from its constituents with SDS–polyacrylamide gel electrophoresis (PAGE). *In situ* digestion was performed on the bands corresponding to the complex and the unbound form of the porin, followed by MALDI MS analysis and identification of the fragment peptides. The proteolytic peptide patterns of the two bands were significantly different, due to the different accessibility of the cleavage sites located in the interface region (40). However, solubilization of membrane proteins with SDS was shown in the past to result in significant structural changes (41); therefore, any conclusions regarding the structure of membrane proteins should be subject to additional scrutiny to avoid possible artifacts.

9.1.3. Analysis of Membrane Proteins Utilizing Other Membrane Mimics

It is a well known fact that strong detergents can denature proteins (17); in fact, Konermann and co-workers (41) recently demonstrated that solubilization of a membrane protein in SDS induces the same conformational changes

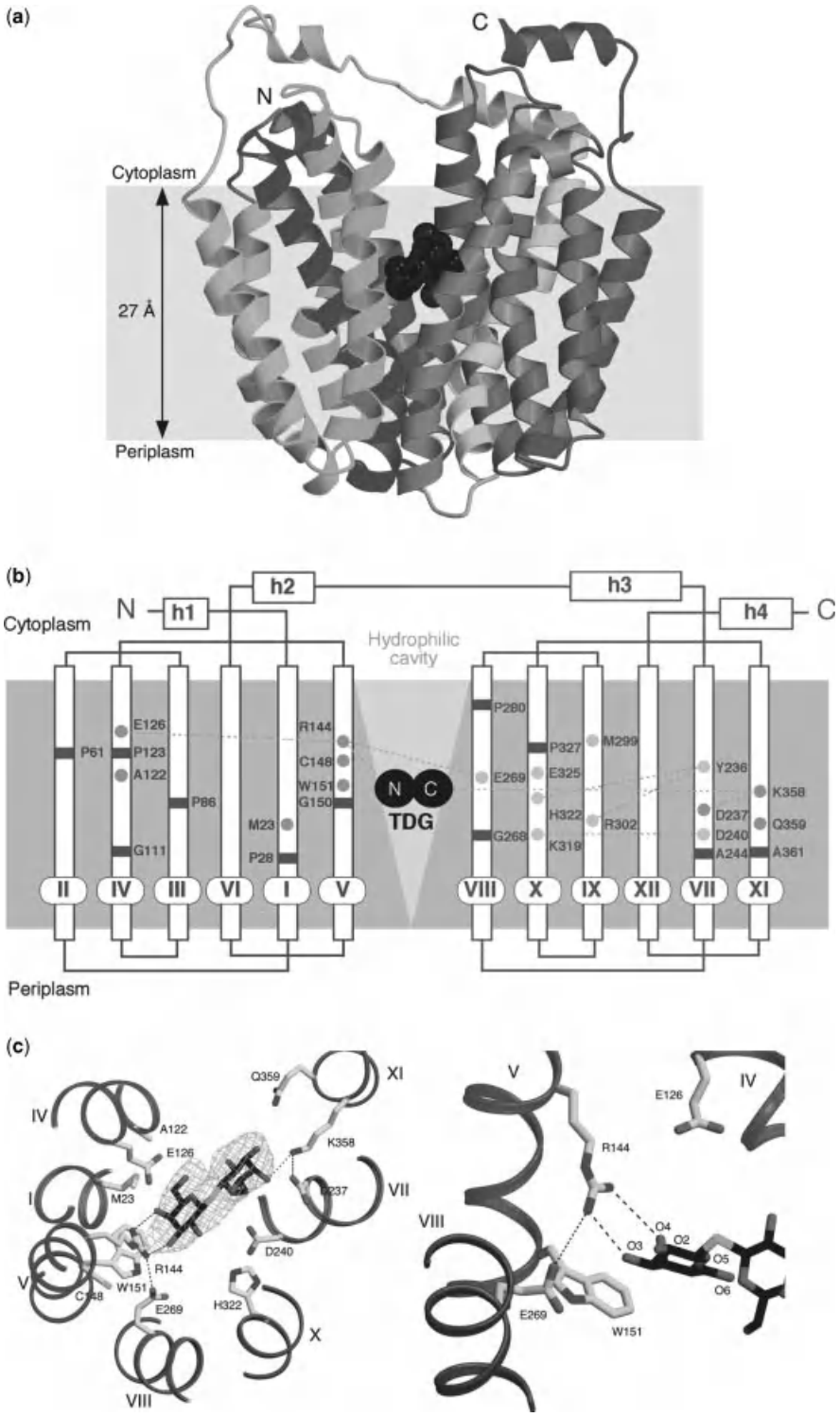


Figure 9.4. Overall structure of lactose permease from *E. coli* (LacY) with a bound substrate homologue TDG. (*a*) Ribbon representation of LacY viewed parallel to the membrane. The 12 transmembrane helices colored from the N-terminus in purple to the C-terminus in pink, and TDG is represented by black spheres. (*b*) Secondary structure schematic of LacY with the N- and C-terminal domains of the transporter colored blue and red, respectively. Residues marked with green and yellow circles are involved in substrate binding and proton translocation, respectively; residue Glu^{269} (light blue circle) is involved in both substrate binding and proton transfer. The hydrophilic cavity is designated by a light blue triangle; h1 to h4 denote surface helices. (*c*) Substrate-binding site of LacY (transmembrane helices in the N- and C-terminal domains are colored blue and red, respectively). [Reproduced with permission from (38).] (See color version of the figure in Color Plate section)

that are seen upon its thermal denaturation at 100°C ! Much less appreciated is the fact that even mild detergents may influence the properties of membrane proteins to a great extent. In Section 9.1.2, we mentioned one example where the choice of a detergent affected the outcome of the stoichiometric measurements for a membrane protein complex (34). There are several well-documented examples where complete solubilization of a membrane protein cannot be achieved without loss of activity, even if the detergents are mild and do not grossly affect the conformation of the protein after solubilization (16). This is hardly surprising, since the behavior of a solubilized membrane protein should be modulated by its interactions with the detergent molecules, although the exact mechanism of protein–detergent interactions remains the subject of intense debate. For example, it is still unclear whether solubilization of the hydrophobic segments of membrane proteins proceeds via their being engulfed by micellar-like structures or by forming a detergent monolayer on the protein surface. Regardless of the specific mechanisms by which the properties of detergent-stabilized membrane proteins are influenced by the detergent molecules, the emerging consensus is that detergent micelles never fully mimic a biological membrane (15,42).

With few exceptions, biological membranes consist of phospholipid bilayers, which may also contain cholesterol. The lipid composition of membranes is not only organism- and organelle-specific, but varies across different cell types (15). An ideal membrane mimetic would be the one that not only forms a bilayer-based structure, but also reflects physical properties of the specific biological membrane it is designed to emulate. There is a wide variety of such structures, including large and small unilamellar vesicles, multilamellar vesicles, amphipols, and nanodiscs; a detailed description of these membrane mimics can be found in an excellent recent review (15). Several experimental approaches are currently being used to probe the structure and behavior of membrane proteins using bilayer-based membrane mimics and MS based methods of detection. One group includes an array of methods that use proteolytic degradation to identify membrane-bound protein segments. Another utilizes hydrophobic probes to obtain topological information on such segments. Finally, HDX can be used to provide information on interfacial positioning and stability of transmembrane polypeptides in lipid bilayers. These methods will be discussed briefly in this section.

Producing peptide maps of membrane-bound proteins may yield information on their topological arrangement by identifying the segments that are confined to a membrane and, therefore, are protected from proteolytic enzymes. Wu and Yates (6) suggested that proteinase K, a relatively nonspecific enzyme, can be used to cleave selectively soluble domains of membrane proteins, followed by MS assisted identification of the fragment peptides. Proteolysis of the "solvent-exposed" soluble domains can be temporally separated from the "protected" segments of these domains, providing further structural differentiation. Smith and co-workers (43) described another method of analysis of membrane proteins that specifically targets the transmembrane segments. Knapp and co-workers (44) used site-specific cleavages of the cytoplasm-exposed loops of rhodopsin induced upon activation of a Cu–phenanthroline tethered cleavage reagent attached to the protein. The cleavage site was identified by mass-analyzing the cyanogen bromide cleaved fragments of the protein, which provides an unbiased mapping of rhodopsin (45).

Selective chemical modification is perhaps one of the most popular techniques currently used to probe the structural arrangements of various domains of membrane-bound proteins. An array of existing hydrophobic probes would only modify protein segments confined to the membrane, while hydrophilic probes would only label solvent-exposed regions of the protein. Hydrophobic photoreactive probes [particularly benzophenone photophores (46)] have been particularly popular in recent years. Benzophenone (BP)-containing photoreagents can be manipulated at ambient light. Activation of BP based photoprobes (by irradiation at 350–360 nm) does not cause protein damage (other than selective chemical modification), while providing a means to control the extent of covalent modification. Importantly, BP based photoprobes react preferentially with inert C−H bonds (Fig. 9.5), with the following reactivity order: $NCH_x > SCH_x >$ methine > $C{=}CCH_2 > CH_2 > CH_3$. Leite et al. (47) presented a particularly intriguing example of how such photoreagents can be used to characterize structural changes within the membrane protein induced by physiologically relevant variations of the physical parameters of the membrane. Several hydrophobic photoreactive agents (Fig. 9.5) were used to characterize the differential accessibility of the nicotinic acetylcholine receptor α_1 subunit in the open, closed, and desensitized states. Photoactivation of the probes was carried out by ultraviolet (UV) irradiation during pulses of the voltage across the membrane ranging from +40 (producing closed states) to −140 mV (producing an $\sim$ 1:1 mixture of open and closed states). Labeling defined the lipid-exposed parts of the transmembrane segments of the protein. More importantly, the results of such experiments helped to identify protein segments that are involved in gating-dependent conformational shifts (47).

The use of HDX MS to map transmembrane regions and characterizing their dynamics using amide hydrogen exchange was pioneered by Heck and co-workers (48,49). Short-model transmembrane peptides were reconstituted in fully hydrated dispersed phospholipid bilayers, and the entire mixture was analyzed by nano-ESI MS. The ESI spectra of the suspension of vesicles clearly showed that the lipid bilayer structures decomposed once they were

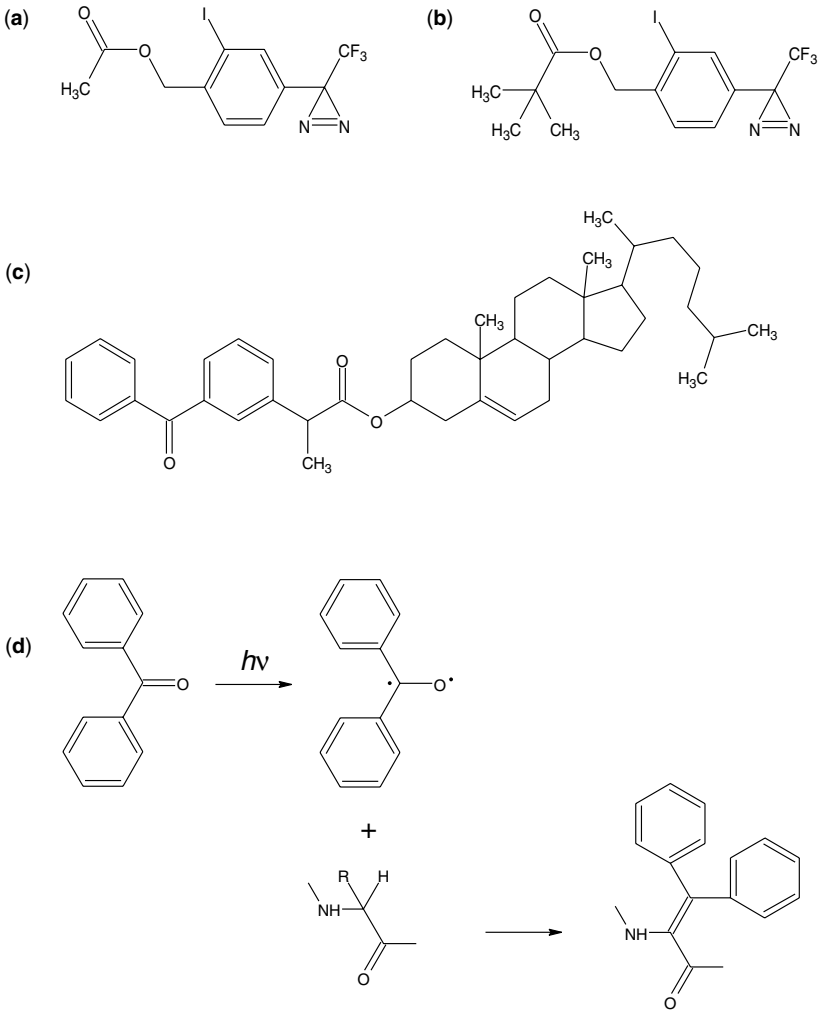


Figure 9.5. Chemical structures of hydrophobic photoreactive probes used in (47) to probe topology of a membrane protein: acetate (*a*), and trimethyl-acetate (*b*) derivatives of 3-trifluoromethyl-3-*meta*-iodophenyl diazirine; and cholesteryl benzoylphenyl propionate (*c*). A schematic diagram depicting the mechanism of photolabeling by benzophenone is depicted in (*d*).

removed from the aqueous environment and transferred to the gas phase. This result should not be surprising in light of our earlier discussion in Section 4.1 of the stability of various types of noncovalent associations in a solvent-free environment. Hydrogen exchange was initiated by diluting the suspensions in deuterated ammonium acetate buffer and the progress of HDX was measured by monitoring the mass gain of the peptide ions. Approximately 10 hydrogen atoms remained protected against exchange, indicating effective solvent shielding by the lipid bilayer from the $(LeuAla)_5$ repeat. The length of this repeat is close to the hydrophobic thickness of the dimyristoylphosphatidylcholine (DMPC) bilayer of $\sim$23 Å in the fluid state, which suggests that this hydrophobic segment is completely embedded in the hydrophobic region of the bilayer. The combination of HDX in solution and peptide ion fragmentation in the gas phase provided an even more definitive proof that this repeat is embedded in the hydrophobic portion of the membrane and is effectively shielded from the solvent. The local deuterium content levels measured as mass gains clearly indicated that amide exchange within the central region of the peptide is extremely slow compared to the terminal (solvent-exposed) segments. A detailed analysis of the local exchange patterns showed some (rather slow) deuterium incorporation at the ends of the transmembrane segment (48), which can be attributed to limited hydrogen scrambling within the peptide ions prior to their fragmentation (50). Alternatively, local dynamic fraying can be invoked to explain such behavior. Indeed, lipid bilayers are known to be dynamic systems and to allow significant water penetration into the hydrophobic interior (51). Hydrogen exchange in the transmembrane segments due to such penetration would be slow for several reasons. The penetration events are rather rare (51) and the hydrophobic environment may impose

severe steric constraints on the water molecule. Furthermore, the dielectric constant in the membrane interior is much lower than that of bulk water itself, which would also adversely affect the exchange rate.

A similar approach was used to explore the effects of transmembrane segment length and composition on local protection patterns (50). One particularly remarkable conclusion of this work was that the insertion of a proline residue in the middle of the transmembrane segment resulted in a significant increase of the exchange rates, while incorporation of a glycine residue did not have any noticeable effect on the stability of the membrane-bound segment of the peptide. Importantly, site-specific measurements of deuterium incorporation into proline-containing membrane-bound peptides clearly show increased conformational flexibility of the transmembrane segments and enhanced water penetration (50).

A very interesting study was reported by Broadhurst and co-workers (52), who used HDX MS to probe the properties of a transmembrane fragment of the M2 protein of Influenza A incorporated into lipid vesicles and detergent micelles. Interestingly, the HDX rates of this protein were significantly lower when the peptide was incorporated into aqueous DMPC vesicles, as compared to those of the peptide in the presence of a large excess of detergent (Triton X-100). In turn, the HDX rate of the peptide incorporated into detergent micelles was significantly lower than that of the denatured peptide in methanol. The results of this study emphasize that the dynamics and solvent accessibility of the transmembrane segments are greatly affected by the properties of the membrane mimetics (52). Another important observation reported by the authors relates to the optimal sample preparation procedure for direct ESI MS analysis of lipid vesicle-bound peptides. Electron microscopy experiments provided evidence that concentration of vesicle by centrifugation induced aggregation, leading to the formation of large multilamellar aggregates, which could not be analyzed by ESI MS (Fig. 9.6). However, the liposome size profile could be maintained when lyophilization was used, followed by thawing above the liquid crystal transition temperature of the lipid component (52).

A recently introduced bilayer-based membrane-mimicking system is a nanodisc, where the bilayer structures are maintained by membrane scaffold proteins modeled after apolipoprotein A1. The utility of this system for HDX MS studies of membrane proteins was recently demonstrated by Engen and co-workers (53), who adapted the HDX MS workflow to incorporate nanodiscs (Fig. 9.7). Another interesting recent example of HDX MS characterization of membrane proteins was recently reported by Woods and co-workers (54), who used this technique to study the interaction of lipoprotein-associated phospholipase A2 with phospholipid (DMPC) vesicles. The studies briefly summarized in this and preceding paragraphs clearly demonstrate the great potential of HDX MS experiments as a tool to probe both topology and dynamics of membrane-bound polypeptides. It remains to be seen if these methods can be applied to larger systems, such as integral membrane proteins. Although there has been significant progress in understanding the mechanisms of folding of water-soluble proteins, the factors governing correct folding and dynamic behavior of integral membrane proteins remain largely unknown (55). Information on the dynamics of these proteins within their native environment would certainly be indispensable for providing answers to an array of important fundamental questions related to "lipid-assisted" protein folding.

Membranes themselves are not static structures, and their dynamic properties are greatly affected by their protein components (56). An important question that still awaits an answer is whether MS can be useful in understanding how proteins influence the mechanical properties of membranes at the molecular level. Deciphering the mechanisms of such processes will obviously have a profound effect on our understanding of a variety of biological phenomena ranging from endocytosis (57) to cell lysis (58–60) to viral entry into the host cell (61,62). An intriguing example of using HDX collision-activated dissociation (CAD) MS to

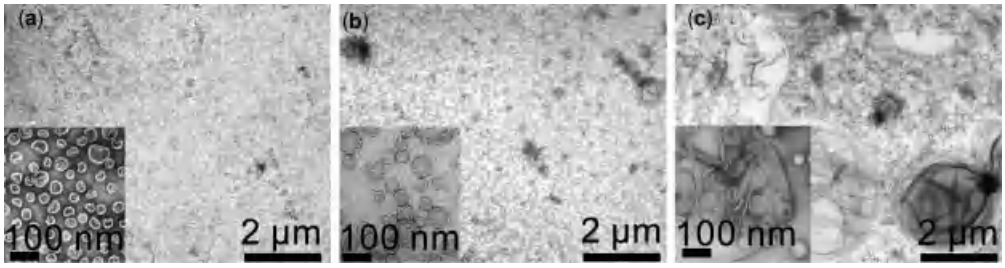


Figure 9.6. Images of mixtures of DMPC liposomes incorporating hydrophobic peptides prepared (*a*) by dialysis of mixed micelles containing the peptide, phospholipid, and the detergent; (*b*) concentrated by lyophilization after the dialysis; and (*c*) concentrated by centrifugation after the dialysis. [Reprinted with permission from (52). Copyright © 2002 American Society for Mass Spectrometry.]

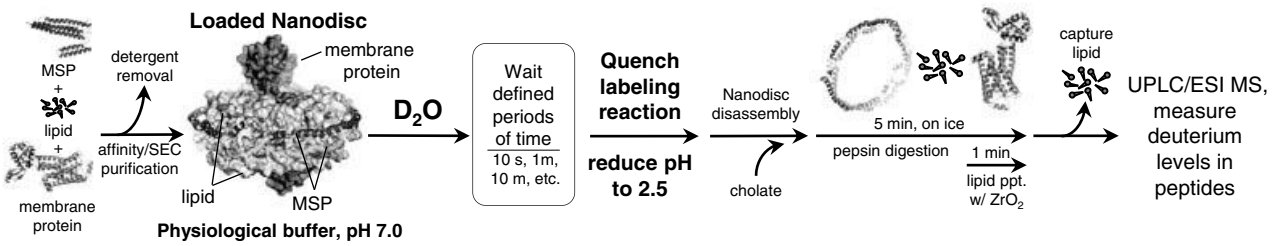


Figure 9.7. The integrated HDX MS workflow on membrane proteins inserted in nanodiscs. Loaded nanodiscs are assembled from a mixture of membrane scaffold protein, lipids, and the target membrane protein solubilized with a detergent. Nanodiscs self-assemble as detergent is removed. Loaded nanodiscs are purified with size exclusion chromatography (SEC) and exposed to deuterated buffer for various times before quenching the exchange reaction. Cholate is immediately added to the quenched reaction to begin disassembly of the nanodiscs. Protein is digested with pepsin for 5 min on ice. In the last minute of digestion, ZrO_2 resin is added to the digestion mixture to selectively remove the phospholipids. Filtration removes immobilized-pepsin beads and the ZrO_2 resin. Ultraperformance Liquid Chromatography (UPLC)/ESI MS is used to measure the incorporation of deuterium. [Reprinted with permission from (53). Copyright © 2010 American Chemical Society.]

study polypeptide interaction with a phospholipid micelle was reported by Akashi and Takio (63). Measurements of the amide protection of melittin, a peptide from bee venom with hemolytic activity, was carried out in the absence and presence of micelles formed by DPC. Although amide exchange of melittin in aqueous solution was very fast, the presence of DPC micelles resulted in a significant decrease in the exchange rate. Site-specific measurements of deuterium incorporation (carried out using collisional activation of melittin ions in the ESI interface) indicated that only one short segment of the peptide, located next to the helix-breaking proline residue, remained relatively flexible. It remains to be seen if this feature is related to the membrane-lytic properties of melittin, as studies by Dempsey and co-workers (64) indicate that lipid perturbations induced by monomeric melittin are very modest, unlike those caused by melittin dimers under identical conditions. It appears that a carefully crafted experimental strategy that involves both HDX and chemical cross-linking may actually provide a detailed account of the melittin–membrane interaction.

9.1.4. Analysis of Membrane Proteins in Their Native Environment

It is difficult to argue with the assertion that the best environment to study the behavior of a membrane protein is in its native membrane. Konermann and co-workers (65) used this philosophy to probe conformational properties of bacteriorhodopsin (BR) in its natural lipid bilayer, purple membranes. Purple membrane suspensions were exposed to hydroxyl radicals, generated by nanosecond laser photolysis of dilute aqueous H_2O_2. Hydroxyl-labeled BR molecules were analyzed by tryptic peptide mapping and ESI MS/MS. Oxidative labeling of BR was found to occur only at its nine methionine residues, and the extent of modification of each of them was consistent with expectations based on the native structure of this protein. Native membranes were also used by Gross and co-workers (66) to study subunit organization of a large membrane complex involved in photosynthesis by mapping (via selective chemical labeling) solvent-exposed side chains of glutamic and aspartic acid residues.

Zhu and co-workers (67) recently reported an impressive example of probing the structure and behavior of membrane proteins in their native environments, who used an hydroxyl radical footprinting to elucidate structural dynamics of the integral membrane porin OmpF in live *E. coli* cells. The cells were exposed to $HO^{\bullet}$ radicals generated by an *in situ* reaction between Hydrogen Peroxide (H_2O_2) and ethylenediaminetetracetic acid (EDTA)-bound Fe^{2+}. Following $HO^{\bullet}$ attack, the outer-membrane proteins were isolated and enriched from the cell lysate. The protein was subsequently purified by SDS–PAGE followed by proteolysis and LC–MS/MS identification and analysis of modified petides. The resulting structural information was in agreement with the available X-ray crystal structure, with reproducibly oxidized amino acid residues being localized in external loops, transmembrane β-strands, and periplasmic turns of the protein (67). We briefly return to a discussion of this work in Section 9.5.2.

9.2. THE PROTEIN AGGREGATION PROBLEM

9.2.1. The Importance and Challenges of Protein Aggregation

Aggregation is arguably one of the oldest phenomena among those ever encountered in protein chemistry and biophysics, although initially it was considered primarily from a utilitarian point of view, that is, either as a means of

protein separation and isolation, or as a tremendous nuisance that seemed to permeate every field of protein research. This determined in many ways the focus of the early studies of this phenomenon (68). It was not until much later that the intimate involvement of aggregation processes in the etiology of *conformational diseases* was realized, giving a fresh impetus and renewed sense of urgency to studies aimed at understanding the mechanism of protein aggregation and amyloid fibril formation. More recently, protein aggregation has come to light in the burgeoning field of biotechnology, particularly in the biopharmaceutical sector, due to the severely deleterious effects of aggregation on the quality and, indeed, safety of protein therapeutics.

Aggregation processes in general are incredibly complex; even the term "aggregate" can mean different things depending on the context, and can encompass structures as diverse as amorphous aggregates, soluble oligomers, amyloid fibrils, and so on (69). Protein energy landscapes, discussed in Chapter 1, offer arguably the best way to visualize the complexity of aggregation and interconnection among different species that can be populated in the course of this process (Fig. 9.8). Until very recently, few experimental approaches were found useful to study aggregation, and the scope of inquiry was usually limited to studies of the endpoint of the aggregation process, while very little information was available on the structure and composition of any intermediate states. As will be seen in the following sections, MS based methods of analysis are now increasingly being used to fill this gap, frequently providing unique information that is critical for understanding of the aggregation phenomenon.

9.2.2. Direct Monitoring of Protein Aggregation and Amyloidosis with Mass Spectrometry

It is not unusual to see soluble protein oligomers in ESI MS, especially if the measurements are carried out at elevated protein concentrations, and it is not uncommon to hear an argument that these species may represent the early stages of protein aggregation. However, this point of view is not universally accepted and, in fact, these oligomeric species are frequently viewed as nonspecific byproducts of the ESI process. Regardless of the origins of these species, even when the initial steps of the aggregation process can be observed directly using ESI MS (70), monitoring the progress of aggregation is far from straightforward.

Miranker and co-workers (71) designed a method to monitor fibrillogenesis in a quantitative fashion and applied it to characterize the aggregation of islet amyloid polypeptide (IAPP), an amyloidogenic peptide. This approach, based on ESI, is complementary to existing assays of fibril formation as it directly monitors the population of precursor rather than product molecules. Fiber formation is monitored in two modes: a *quenched* mode in which fibril formation is halted by dilution into denaturant, and a *real-time* mode in which fibril formation is conducted within the capillary of the ESI source. The fibrillar IAPP does not compromise the ionization of the monomeric form of the peptide.

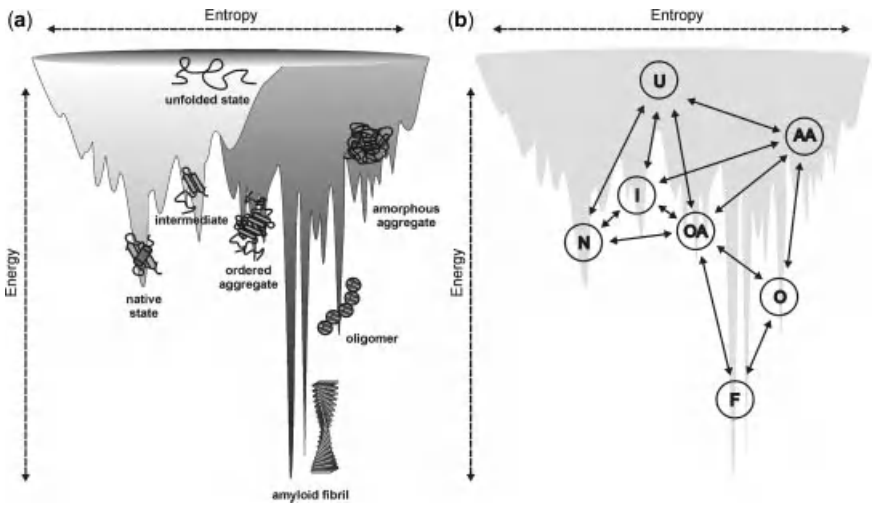


Figure 9.8. Illustration of a combined energy landscape for protein folding and aggregation. (*a*) the surface illustrates the roughness of the protein energy landscape, showing the multitude of conformational states available to a polypeptide chain. While rather simple folding funnels (light gray) can describe the conformational search of a single polypeptide chain to a functional monomer, intermolecular protein association dramatically increases ruggedness (dark gray). (*b*) proposed pathways linking the conformational states shown in (*a*) populated on the combined folding and aggregation energy landscape. [Adapted from (69). Copyright © 2008 with permission from Elsevier.]

Furthermore, under mild ionization conditions, fibrillar IAPP does not dissociate and subsequently contribute to the monomeric signal. Quantitation was made possible through the introduction of an internal standard, IAPP from another animal species. The standard (rat IAPP) is structurally close to human IAPP, so that the ionization efficiencies of the two peptides are identical. At the same time, the masses of the two peptides are slightly different, so that they do not interfere with each other in the ESI mass spectrum. When applied to IAPP fibrillogenesis, the technique revealed that precursor consumption in seeded reactions obeys first-order kinetics, while a consistent level of monomer persists in both seeded and unseeded experiments after fibril formation is complete (71).

Another MS based procedure to monitor conformational changes that precede the formation of fibrils was developed by Serrano and co-workers (72), who tracked aggregation of a model protein ADA2h (activation domain of human procarboxypeptidase A2) by monitoring the disappearance of protein monomers and generation of resistance to proteolysis. Limited proteolysis of amyloid fibrils monitored by MS was also employed by Pucci and co-workers (73), who studied the structural determinants of amyloid nucleation. Seven proteases, with different degrees of specificity, were used to map solvent-exposed regions within the native, ligand-bound and amyloidogenic states of muscle acylphosphatase (AcP), a protein that forms amyloid fibrils in the presence of TFE. The results of this study suggest that a considerable degree of solvent exposure is an important structural feature that initiates the protein aggregation process (73). Limited proteolysis monitored by MALDI or ESI MS has been used to characterize the structure of amyloid fibrils in a number of other studies (74,75). Monitoring the rate of disappearance of protein monomers with ESI MS as a means of evaluating aggregation kinetics was employed more recently by Grubmüller and co-workers (76) for insulin, and Andrisano and co-workers (77) for Aβ(1–42) peptide. The authors of the latter work also noted the presence of low molecular weight aggregates in MALDI MS, which were suggested to represent early aggregation nuclei; however, the relevance of such species to the aggregation process remains unclear. A very interesting example of monitoring early oligomers of amyloidogenic β2-microglobulin with ESI MS was presented by Radford and co-workers (78), who were able to detect and identify transient aggregation, highlighting the unique capability of this technique to characterize individual species formed early during amyloid assembly.

In the biopharmaceutical sector, progression of protein aggregation is commonly probed with SEC, (discussed in Section 2.5.4). While SEC, in many cases, allows various low molecular weight oligomers to be resolved, or at least their heterogeneity to be visualized, it cannot identify these species without input from other techniques. Mass spectrometry (ESI MS in particular) has been used in the past as a means of identifying SEC fractionated molecules, and has even been incorporated into an online SEC–MS scheme (79), although its applications are limited mostly to synthetic polymer analysis. Recently, Heck and co-workers (80) used off-line mass spectrometric analysis of SEC fractions of acid-stressed monoclonal antibody (IgG) to characterize intact soluble oligomers. Oligomer fractions were isolated and dialyzed against ammonium acetate, pH 6.0, followed by ESI MS analysis, which identified several low molecular weight aggregates (up to tetramers). While ESI MS was shown to leave the IgG oligomers largely intact, some dissociation was noticeable within the collected fractions prior to MS analysis (80). Recently we used a similar approach to characterize soluble aggregates of recombinant antithrombin III (AT-III) formed under heat-stress conditions (81). To reduce the possibility of post-SEC alteration of fraction composition observed in (80), ESI MS measurements were carried out directly on SEC fractions without any postcolumn processing. The resulting mass spectra of SEC fractions are apparently unaffected by oligomer dissociation, and allow the fractions to be easily identified up to tetramers (Fig. 9.9). The presence of multiple oligomeric species in higher fractions, as well as structural heterogeneity of AT-III, makes their confident identification difficult, although it might be possible that this problem will be resolved in the future using the complexity reduction techniques (82) discussed in Chapter 4, as well as later in Section 9.3.

A very interesting question that arises in connection with the SEC chromatogram shown in Figure 9.9 is how the abundance of individual oligomers correlates with the ionic signal intensity of such species in ESI MS. Conventional wisdom (to which the authors of this book subscribed until very recently) holds that ESI MS always introduces a bias against higher molecular weight proteins. Therefore, any attempt to study the composition of a mixture of soluble oligomers would inevitably exaggerate the contribution of monomers and diminish that of higher oligomers. A quick comparison of the ESI mass spectrum of unfractionated heat-stressed AT-III (Fig. 9.10) and its SEC chromatogram seem to confirm this suspicion. However, we need to keep in mind that the SEC data reflects mass concentration of oligomers (as it relies on UV detection), while ESI MS reflects molar concentration (with a single ion producing a single count regardless of its size). In other words, in the absence of any biases, a tetramer molecule would generate four times the signal of the monomer in SEC, while a tetramer ion will generate the same signal as a monomer ion in ESI MS. Once this correction is made, the distribution of oligomeric ions in ESI MS (inset in Fig. 9.10) shows remarkable similarity to the distribution of oligomers provided by SEC (Gaussian peaks in Fig. 9.9). Therefore, it appears that, at least in some cases, ESI MS can be applied

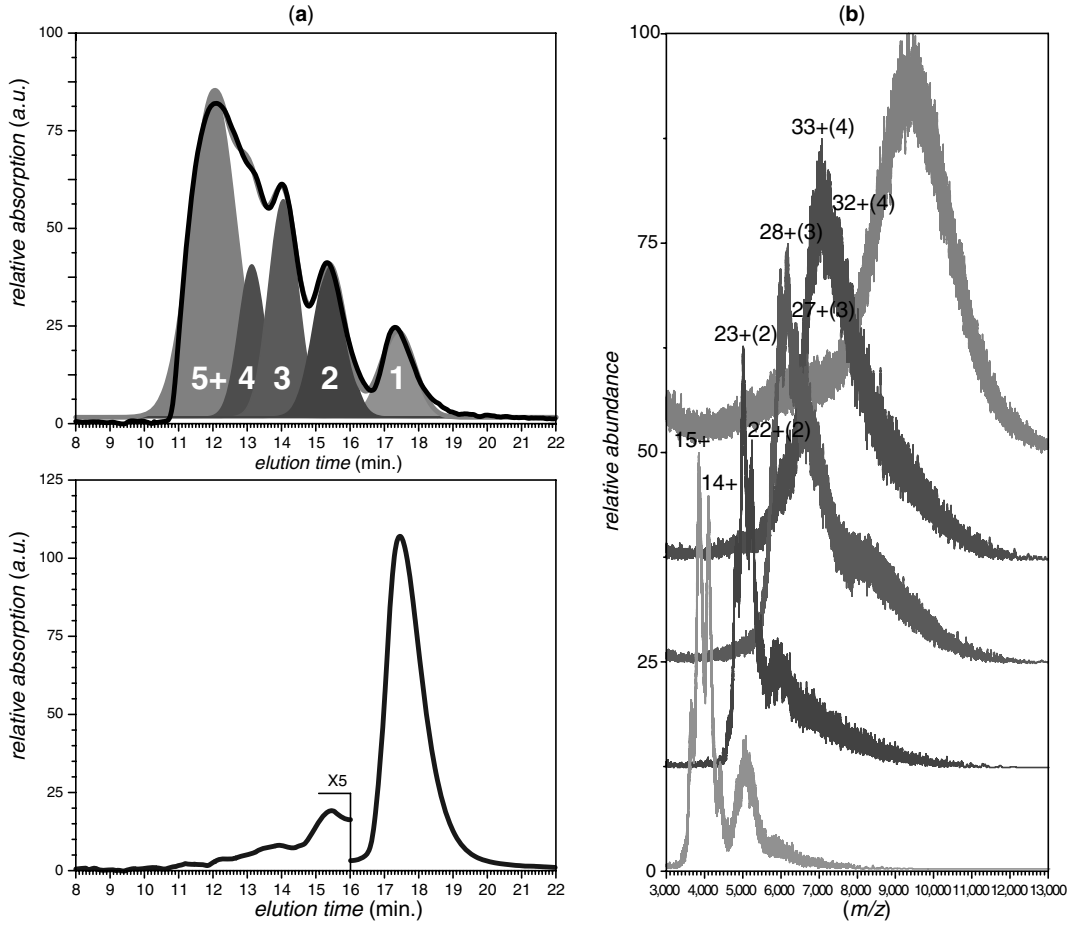


Figure 9.9. SEC and ESI MS analysis of soluble aggregates of AT-III produced by heat stress. The chromatograms in (*a*) show SEC profiles of stressed (top) and untreated (bottom) samples. The chromatogram of the stressed sample was fitted with several Gaussian curves, which were then used to determine optimal fraction collection times. The ESI MS analysis of each fraction's content (*b*) allowed the identities of soluble oligomers to be established up to a tetramer. [Data courtesy of Guanbo Wang (UMass-Amherst).]

directly to monitor the distribution of soluble aggregates in protein solutions.

While the ability to bypass the separation step discussed in the preceding paragraph does not offer any significant advantages when analyzing static samples, it may become critically important in the analysis of oligomers that become populated transiently during the aggregation process and, therefore, may easily evade detection by SEC. This is best illustrated by the results of our recent studies of heat-induced aggregation of several biopharmaceutical products, where a temperature-controlled ESI source was used as an alternative to accelerated stability studies (83). In this work, a protein solution was heated at a fixed temperature (controlled to within $\pm$ 2°C) for a fixed period of time (ranging from tens of seconds to several minutes) inside the ESI capillary of a mass spectrometer. Minimization of heat loss during the final stages of the ESI process allowed both conformational changes and oligomer formation to be monitored simultaneously as a function of solution temperature and the duration of heat exposure.

We used this approach to monitor heat-induced aggregation of the 63 kDa glycoprotein human glucocebrosidase (GCase), a biopharmaceutical product that is used to treat Gaucher's disease. Although heat stress is frequently used to test the stability of protein drugs, classical biophysical techniques that are commonly employed to detect the onset of destabilization in response to elevated temperature do not provide detailed description of this process at the molecular level. For example, the DSC profile of GCase shows a transition at 49°C, and the shape of the peak suggests that the transition is irreversible. Beyond that, it reveals very little about the nature of the underlying processes. On the other hand, examination of the behavior of heat-stressed GCase with temperature-controlled ESI MS (Fig. 9.11) provides very rich information content (83). Indeed, not only does it allow the onset of aggregation to be detected,

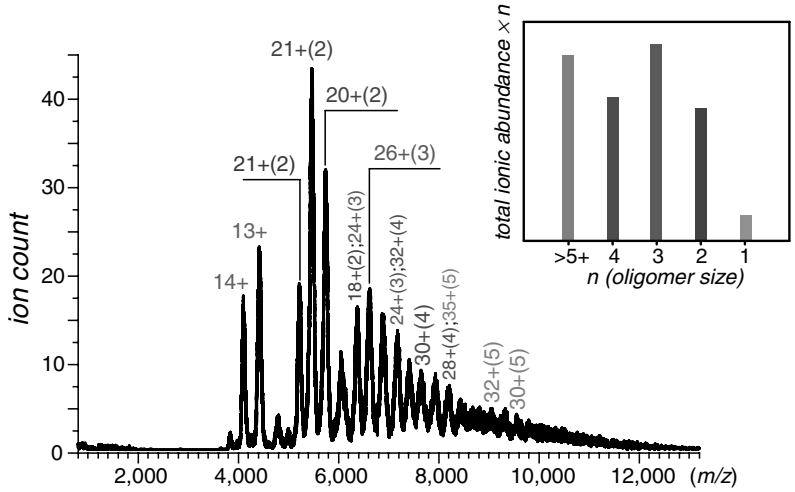


Figure 9.10. The ESI MS of unfractioned soluble aggregates of AT-III produced by heat stress. The inset shows total ionic abundance of each identified oligomer recalculated to reflect mass concentration (i.e., the total ionic abundance of each oligomer was multiplied by the number of monomeric units n). [Data courtesy of Guanbo Wang (UMass-Amherst).]

but also yields detailed information regarding specific molecular events that accompany this transition. While changes in the extent of multiple charging of protein ions are relatively minor in the 25–46°C range, further increase of the protein solution temperature clearly leads to a large-scale loss of structure, as evidenced by dramatic changes in the protein ion charge-state distribution (i.e., appearance of the high charge-density protein ion peaks centered around charge state 27+ in Fig. 9.11). This loss of structure at the level of GCase monomers coincides with the onset of GCase oligomerization. Although the extent of oligomerization observed at 50°C is relatively modest, further temperature increase results in progressive growth of their sizes and abundance. Thus, temperature-controlled ESI MS provides a unique ability to monitor the emergence and evolution of soluble protein oligomers, which are likely to play important roles in aggregation. Importantly, the duration of the heat stress in these experiments can be easily changed simply by varying the protein solution flow rate, thereby allowing time-resolved studies of aggregation processes to be carried out (83).

Detailed analysis of GCase behavior as a function of both solution temperature and heat exposure time suggests that protein unfolding is the initial step leading to large-scale oligomerization of heat-stressed GCase, that is, unfolded monomers are obligatory intermediates *en route* to aggregation of this protein (83). Intriguingly, this is not a general rule; in fact, heat-induced aggregation of some proteins appears to be triggered by relatively small-scale conformational transitions that do not result in a global loss of structure. An example of such behavior revealed by temperature-controlled ESI MS measurements is shown in Figure 9.12, where behavior of AT-III was tracked by DSC and ESI MS over a broad temperature range (84). No large-scale unfolding is observed in ESI mass spectra at or even above 56°C (the transition temperature deduced from DSC measurements). However, one can clearly see relatively small changes in the average charge of ions representing compact AT-III species in solution. As already seen in Section 5.1, such changes might be indicative of small-scale conformational transitions that do not result in significant loss of higher order structure. Plotting the average charge of AT-III ions as a function of temperature gives a sigmoidal curve with a midpoint at 55.5°C (Fig. 9.12), which is within the experimental error of the DSC transition! Importantly, the appearance of soluble oligomers in the ESI mass spectra can also be seen at 55°C and higher (Fig. 9.12), further validating the notion of a small-scale conformational transition in AT-III as an event triggering aggregation (84).

9.2.3. Structure of Protein Aggregates, Amyloids, and Pre-Amyloid States

Several research groups have used HDX MS to probe the stability and structure of various amyloid formations, as well as pre-amyloid protein states. In a typical experiment, amide protection patterns are acquired for the free protein in solution, and also in the amyloid form. A comparison of the two exchange patterns often provides valuable information on the structural changes that accompany amyloid formation (85,86). Robinson and co-workers (70) employed HDX MS to monitor the backbone protection of insulin molecules within relatively small aggregates in an oligomer-specific fashion. The ESI mass spectra were obtained from solutions of insulin at millimolar concentrations at pH 2.0, conditions known to promote insulin aggregation. Clusters containing up to 12 insulin molecules could be detected in the gas

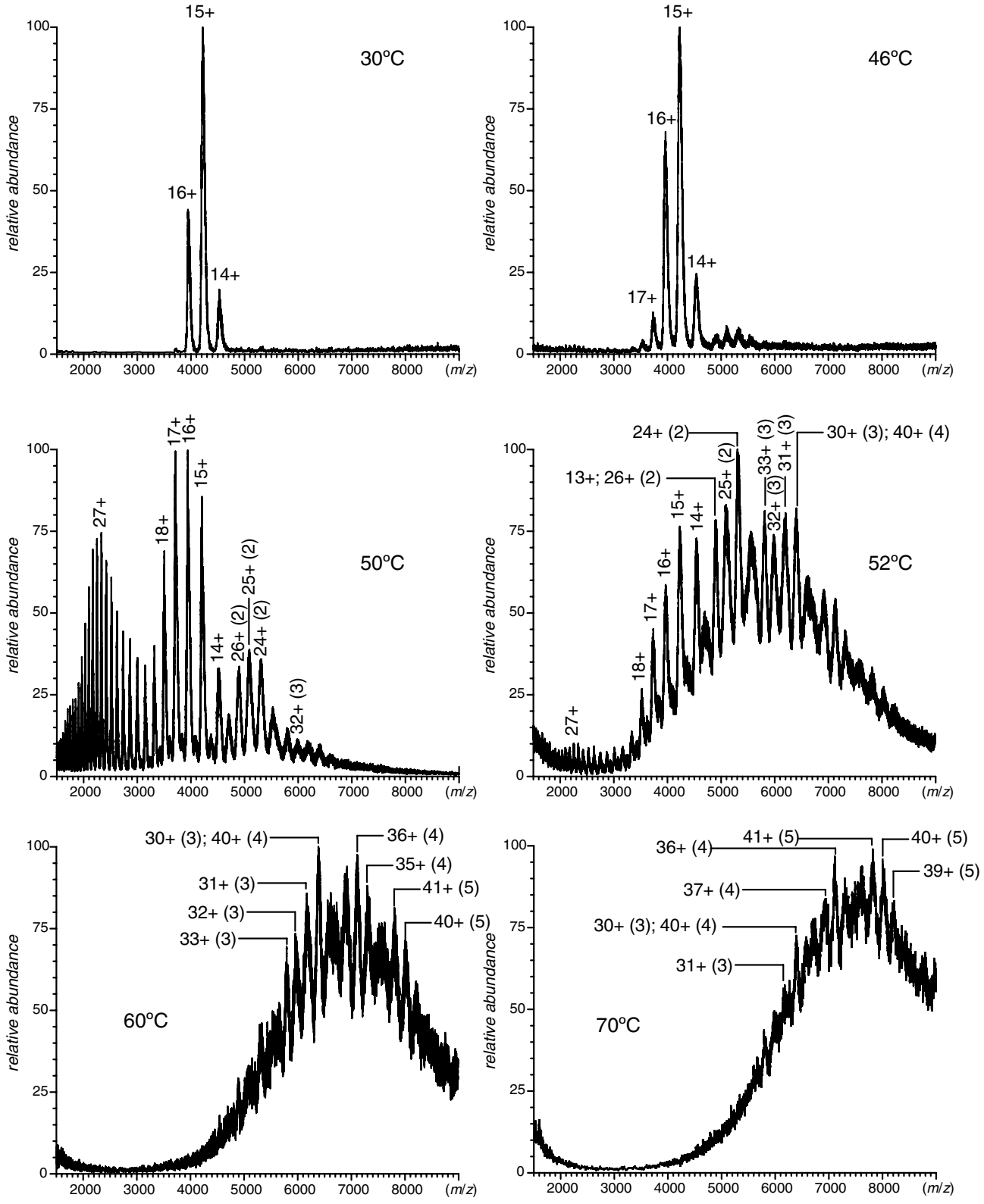


Figure 9.11. The ESI mass spectra of GCase (1 μM in 20 mM ammonium acetate, pH 4.7) recorded at various solution temperatures. The numbers without parentheses indicate charge states of GCase monomers; the numbers in parentheses indicate the size of protein associations. [Adapted with permission from (83). Copyright © 2011 American Chemical Society.]

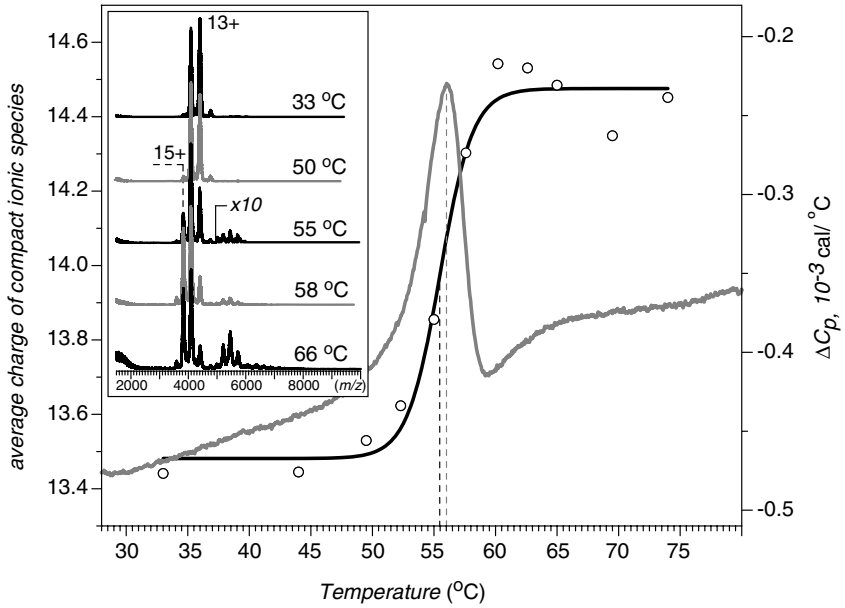


Figure 9.12. Temperature-dependent behavior of AT-III monitored by DSC and ESI MS. Both DSC profile (gray curve) and the average charge of AT-III ions in ESI MS show the same (within the experimental error) transition temperature at 56°C. The inset shows several mass spectra of AT-III acquired at different temperatures. In addition to a small shift of the charge-state distribution of monomer ions (*m/z* 3500–4500), appearance of soluble oligomers (*m/z* > 5000) can be seen at 55°C and higher. [Data courtesy of Guanbo Wang (UMass-Amherst).]

phase. The HDX MS measurements showed that in solution these higher oligomers are in rapid equilibrium with monomeric insulin. At elevated temperatures, under conditions where insulin rapidly forms amyloid fibrils, the concentration of soluble higher oligomers was found to decrease with time yielding insoluble, high molecular weight aggregates, and then fibrils (70). Further studies of insulin aggregation by the Robinson group utilized a combination of HDX and peptic digestion (performed after quenching the exchange) to obtain site-specific information on amide protection within insulin molecules derived from the fibers (87).

Nazarbal et al. (88) used HDX MS to study the conformational transition occurring upon amyloid aggregation of the prion protein HET-s, an infectious element of the filamentous fungus *Podospora anserina.* The HET-s prion protein is a 289 amino acid residue protein that aggregates into amyloid fibers *in vitro*. Such fibers produced *in vitro* are infectious, indicating that the HET-s prion can propagate as a self-perpetuating amyloid aggregate of the HET-s protein. The HDX of the soluble form of HET-s monitored using MALDI MS revealed that the C-terminal region of the protein (240–289) exhibits a high solvent accessibility. However, solvent accessibility is drastically reduced in this region in the amyloid form of the protein. The HDX pattern in the N-terminal part of the protein [residues (1–220)] is not significantly affected by protein aggregation, suggesting that aggregation of HET-s involves a conformational transition only in the C-terminal part of the protein (88).

Lansbury and co-workers (89) used HDX MS to characterize the secondary structure of protofibrils, heterogeneous transient structures observed during the *in vitro* formation of mature amyloid fibrils, which have been implicated as the toxic species responsible for cell dysfunction and neuronal loss in Alzheimer's disease. Populations of protofibrils of different sizes were isolated chromatographically from amyloid assembly reactions of amyloid peptide Aβ (1–40). Hydrogen–deuterium exchange MS was shown to be able to distinguish among unstructured monomer, protofibrils, and fibrils based on their different amide protection patterns. About 40% of the backbone amides of Aβ (1–40) protofibrils were highly resistant to exchange even after 2 days of incubation in aqueous deuterated buffer, implying a very stable core structure. This is in contrast to the mature amyloid fibrils, whose equally stable structure protects ~60% of the backbone amide hydrogen atoms over the same time period. A surprising degree of specificity was observed in amyloid assembly, with the wild-type Aβ (1–40) being preferentially excluded from both protofibrils and fibrils grown from an equimolar mixture of wild-type and mutant peptides (89).

Aggregation properties of Aβ (1–40) were also studied by Fernandez and co-workers (90), who used HDX followed by the proteolytic (peptic) digest of Aβ (1–40) carried out under slow-exchange conditions to probe the structure of amyloid fibrils. These experiments revealed differential local protection patterns within Aβ (1–40), with the N-terminus

[residues (1–4)] being completely unprotected, the C-terminal segment exhibiting intermediate levels of protection, and the fragment containing residues (5–19) of the peptide exhibiting very high protection in the fibril-containing samples (90). More recent studies by the same group focused on measuring the peptide level solvent accessibility of multiple aggregated states of this polypeptide using HDX MS. Consistent with earlier results, a gradual reduction in solvent accessibility was observed, spreading from the C- to the N-terminal region with ever more aggregated states of Aβ (1–40). The protection was observed to begin with reporter peptides (20–34) and (35–40) in low molecular weight oligomers found in fresh samples, culminating with increasing solvent protection of reporter peptide (1–16) in long time-aged fibrillar species (91,92). A very interesting report was published recently by Robinson and co-workers (93), who used HDX MS to demonstrate that the Aβ molecules comprising the fibril continuously dissociate and reassociate, a phenomenon that was termed molecular recycling. Importantly, the recycling was found to occur on biologically relevant time scales, which obviously has important implications for understanding the role of Aβ amyloid fibrils in neurotoxicity.

Prion protein (PrP) is another paradigmatic amyloid-forming polypeptide, which has gained a great deal of notoriety due to its intimate involvement in the etiology of the "mad cow" disease, transmissible spongiform encepalopathy. Surewicz and co-workers (94) used HDX MS to examine the structure of amyloid fibrils formed by the recombinant human PrP segment (90–231), a misfolded form that was reported to be infectious in mice overexpressing normal PrP. Analysis of the HDX pattern allowed the hydrogen-bonded β-sheet core of PrP amyloid to be localized to the C-terminal region, which in the native structure of PrP monomer corresponds to two α-helices connected by a loop. At the same time, no extensive hydrogen bonding (i.e., no significant protection of backbone amide hydrogen atoms) was detected in the N-terminal part of PrP (90–231) fibrils, arguing against the involvement of residues in this region in stable β-structure, and providing long-sought experimentally derived constraints for high-resolution structural models of PrP amyloid fibrils (94). More recently, the structure of PrP fibrils was re-evaluated by Woods and co-workers (95), who used HDX MS to examine behavior of a shorter PrP segment (89–143) bearing a mutation, P101L, implicated in familial prion disease. Two segment of the polypeptide sequence were found to exhibit anomalously high protection in the fibrillar form of PrP (89–143, P101L), indicating their involvement in the fibril core (95).

Finally, a truly amazing example of using HDX MS to obtain fine structural details of amyloid formations was presented by Englander and co-workers (96), who used unprotected and protected backbone amide hydrogen atoms to define the order–disorder boundaries in α-synuclein aggregates at spatial resolution close to the single amino acid level. Soluble α-synuclein monomers underwent HDX at random coil rates, consistent with the notion of it being an intrinsically disordered protein. However, once this protein was assembled to form amyloids, long N- and C-terminal segments remained unprotected (although the N-terminal segment showed some heterogeneity). A continuous middle segment was found to be strongly protected against exchange, a clear sign of a stable hydrogen-bonded cross-β structure. These results suggested that α-synuclein adopts a chain fold similar to that of Aβ (96).

In Chapter 5, we saw that the top-down approach to deducing local protection patterns from HDX MS measurements frequently offers significant advantages over the classical bottom-up scheme. Top-down HDX MS was first utilized in the studies of amyloid-forming proteins by Kraus et al., (97), who used this technique to investigate the behavior of synthetic β-amyloid peptides. In contrast to the C-terminally truncated peptides Aβ (1–40) and Aβ (1–36) showing lack of protection, Aβ (1–42) and the pyroglutamyl peptide pAβ (3–42) exhibited more complex exchange patterns resulting from the formation of β-sheet structured oligomers with 18–20 highly protected amides. Peptide ion fragmentation in the gas phase indicated that the protected region of Aβ (1–42) is located in the central part of the chain [residues (8–23)], confirming the importance of the hydrophobic-binding domain LVFFA for Aβ aggregation (97). Note that the experimental scheme implemented by Kraus et al. (97) did not utilize tandem capabilities of MS (peptide ion fragmentation was carried out in the ionization source prior to mass selection). In Section 5.4.2, we saw that utilization of the MS/MS capability in HDX MS work allows higher order structure and dynamics to be characterized in a conformer-specific fashion (e.g., see Fig. 5.23). In a very exciting recent development, Konermann and co-workers (98) were able to exploit this unique advantage of top-down HDX MS to carry out conformer-specific HDX analysis of Aβ (1–42) oligomers. In this work, intact ions exhibiting specific mass shifts (reflecting the extent of deuterium incorporation) were selected in the gas phase, followed by ECD. Masses of the resulting fragment ions allowed the hydrogen-bonding pattern to be determined in a conformer-specific and spatially resolved fashion, highlighting the similarities between oligomers and mature amyloid fibrils vis-à-vis backbone organization (98). Importantly, the spatial resolution obtained with the top-down HDX MS approach exceeded that of earlier studies of Aβ conformation and dynamics that relied on the classical bottom-up scheme (see above).

Among other exciting recent developments, we should mention work reported by Robinson and co-workers (99), who employed a pulsed-labeling approach (Fig. 9.13) to characterize aggregation of an SH3 domain of the α-subunit of bovine phosphatidylinositol-3′-kinase (PI3-SH3), a pro-

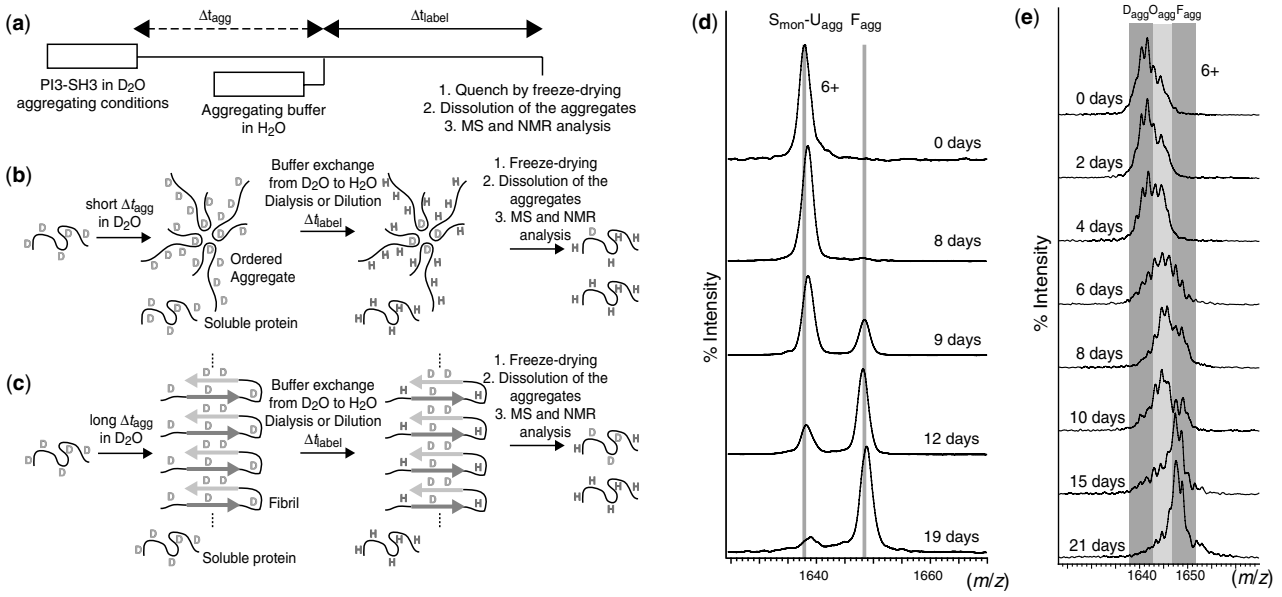


Figure 9.13. Schematic description of the pulse-labeling HDX exchange experiment to study protein aggregation. (*a*) The experiment starts by incubating soluble protein under aggregation conditions in a deuterium based buffer. After a variable aggregation time, Δt_{agg}, labeling takes place for a fixed period, Δt_{label}, using protonated aggregation buffer. The magnitude of Δt_{label} is chosen so that only unprotected amide deuterons will exchange significantly with the solvent. After the labeling pulse, freeze-drying is used to quench exchange. Different samples are prepared at defined Δt_{agg} values, which are later solubilized into monomers by transfer to a DMSO solution and analyzed by NMR and ESI MS. The figure illustrates hypothetical scenarios when the protein is left to aggregate for a short Δt_{agg} (*b*) and a long Δt_{agg} (*c*). (*d*) The HDX MS (6+ charge state) for PI3-SH3 pulse-labeled samples prepared under amorphous conditions. The spectra show the relative populations of S_{mon}-U_{agg} and F_{agg} at the indicated Δt_{agg} times. Peak intensities are normalized to the overall species population. (*e*) The HDX MS (6+ charge state) for PI3-SH3 pulse-labeled samples prepared under protofibril formation conditions. The spectra show the relative populations of D_{agg}, O_{agg}, and F_{agg} at the indicated Δt_{agg} times. Peak intensities are normalized to the overall species population. [Reprinted with permission from (98).]

tein that aggregates to form well-characterized fibrils under acidic conditions *in vitro* (99). Pulsed-labeling HDX MS experiments were carried out under two different conditions, both ultimately leading to well-defined amyloid fibrils. Under one condition (pH 2.0, Fig. 9.13*d*), the intermediates appear to be largely amorphous in nature, whereas under the other condition (pH 1.5, Fig. 9.13*e*) protofibrillar species are clearly evident. Under the amorphous conditions, only species having no protection against hydrogen exchange could be detected in addition to the mature fibrils that show a high degree of protection. By contrast, under protofibril formation conditions, HDX MS reveals that multiple species are present with different degrees of hydrogen-exchange protection (Fig. 9.13*e*), indicating that aggregation occurs initially through relatively disordered species that subsequently evolve to form ordered aggregates, which in turn eventually lead to amyloid fibrils. Subsequent analysis with NMR provided residue-specific information on the structural reorganizations taking place during aggregation (99), a step that might be replaced in the future with HDX MS/MS measurements discussed in the preceding paragraph.

Other recent examples of using HDX MS to probe conformation and dynamics of amyloidogenic or aggregation-prone proteins, under conditions that attenuate the rate of aggregation, were reported by Ashcroft and co-workers (100), who studied the stabilizing influence of major histocompatibility complex 1 on β2-microglobulin (100); and Fernandez and co-workers (101), who studied the effect of repeated free–thaw cycles on protein stability and aggregation. An interested reader is referred to several excellent reviews on various aspects of using HDX MS technology to study protein aggregation processes (85,86,102).

Although HDX MS appears thus far to be the most popular MS based method to study protein aggregation processes, several interesting studies have been published that utilize other methods, such as chemical labeling and cross-linking. For example, Egnaczyk et al. (103) used photoaffinity cross-

linking to determine structural constraints for Aβ monomer arrangement within amyloid fibrils. A photoreactive Aβ (1–40) was synthesized by substituting Phe at position 4 with L-*p*-benzoyl-phenylalanine (Bpa), and then incorporated into synthetic amyloid fibrils, followed by UV irradiation of the peptide aggregate. SDS–PAGE of dissolved fibrils revealed the light-dependent formation of a covalent Aβ dimer. The MS analysis of the products of enzymatic cleavage of the dimers suggestsed that Bpa^4 was covalently attached to the C-terminal segment of a β-amyloid peptide (tryptic fragment (29–40)). The MS/MS experiments and further chemical modifications of the cross-linked dimer allowed the precise location of the chemical modification to be established as the side chain of Met^{35}. The $Bpa^4 \rightarrow Met^{35}$ intermolecular cross-link is consistent with an antiparallel alignment of β-amyloid peptides within the amyloid fibrils (103).

A very useful method for characterizing oligomeric species that has seen a surge in popularity in recent years is photoinduced cross-linking of unmodified proteins, PICUP (104). The photochemistry of PICUP is based on photooxidation of Ru^{2+} in a tris(bipyridyl)Ru(II) complex ($[Ru(bpy)_3]^{2+}$, a commercially available chemical used in a variety of photochemical reactions) to Ru^{3+} by irradiation with visible light in the presence of an electron acceptor. Giese and co-workers (105) used PICUP as a means of stabilizing transient oligomers of PrP (see above) for subsequent MS analysis. The degree of cross-linking within PrP aggregates was adjustable using varying light intensities. Specific intermolecular cross-linking of scrapie-form prion protein (PrP^{Sc}) molecules was achieved even in crude brain homogenate. Functional studies showed that stabilized aggregates of PrP did not lose their capacity to induce further protein aggregation and cross-linking of PrP^{Sc} did not alter significantly the level of infectivity, indicating that photoinduced covalent linkage of PrP^{Sc} does not destroy surfaces important for prion propagation (105). Another recent example of using PICUP in conjunction with MS detection was reported by Teplow and co-workers (106), who used it to study structural and assembly dynamics of Aβ protein. In this work, PICUP was used in combination with scanning amino acid substitution (by replacing various residues with tyrosine, a PICUP reactive amino acid) as a method for quantitative determination of Aβ oligomerization, an approach dubbed *scanning* PICUP. In contradiction to the conclusions of the HDX MS work discussed earlier in this section, it was determined that the N-terminus is not a benign peptide segment; instead, the Aβ conformational dynamics and assembly appeared to be affected significantly by the competition between the N- and C- termini to form a stable complex with the central hydrophobic cluster (106).

The use of bifunctional cross-linkers in studies of protein aggregation and amyloidogenesis is exemplified by work reported by Requena and co-workers (107), who used BS^3 (see Table 4.2) to gain structural information on PrP purified from the brains of rodents infected with scrapie. The cross-linking reaction yielded dimers, trimers, and higher-order oligomers that were separated by SDS–PAGE. In gel digestion, followed by MS analysis, enabled identification of the interchain pairing that was consistent with an existing model of PrP based on electron crystallographic data (Fig. 9.14). Chemical labeling has also been shown to be a useful tool in providing structural information that can be utilized to generate models of various pre-amyloid intermediate states. For example, Vachet and co-workers (108) successfully used selective chemical labeling to generate models of dimeric (108) and tetrameric (109) intermediates of β2-microglobulin populated *en route* to its *in vitro* aggregation. Selective chemical labeling has also been used by Good and co-workers (92) in kinetic studies of Aβ aggregation by following the time course of alkylation of lysine residues (92).

This section would be incomplete without mentioning IM–MS-based approaches to study behavior of protein aggregates. We already saw in Section 4.1.5 that even though structural changes in protein assemblies upon their transition from solution to the gas phase are inevitable, information deduced from measuring their collisional cross-sections in the gas phase is nonetheless useful, and can be related to the solution-phase structures. This is particularly true in the studies of protein aggregation, where dramatic structural differences among various types of aggregates should lead to significant variations of collisional cross-sections (Fig. 9.15). IM–MS has been successfully used to study Aβ (1–42) monomer and its early aggregation states (111–113), as well as to elucidate the self-assembly mechanism of short fragments of amyloidogenic peptides, where unambiguous evidence was provided for structural transitions leading to fibril formation (110).

9.3. THE MANY FACES OF COMPLEXITY: MASS SPECTROMETRY AND THE PROBLEM OF STRUCTURAL HETEROGENEITY

One of the peculiar features of the aggregation processes discussed in Section 9.2 that makes them so difficult to study is the enormous heterogeneity of the repertoire of species that become populated at its various stages. This heterogeneity is displayed at two levels: size heterogeneity (due to a varying number of monomers comprising individual oligomers) and conformational heterogeneity. Another type of heterogeneity that is frequently encountered in proteins is structural heterogeneity caused by post-translational modifications (PTMs). Nonenzymatic PTMs that accumulate as a result of protein aging or stress-related degradation are obviously random, but even biologically relevant

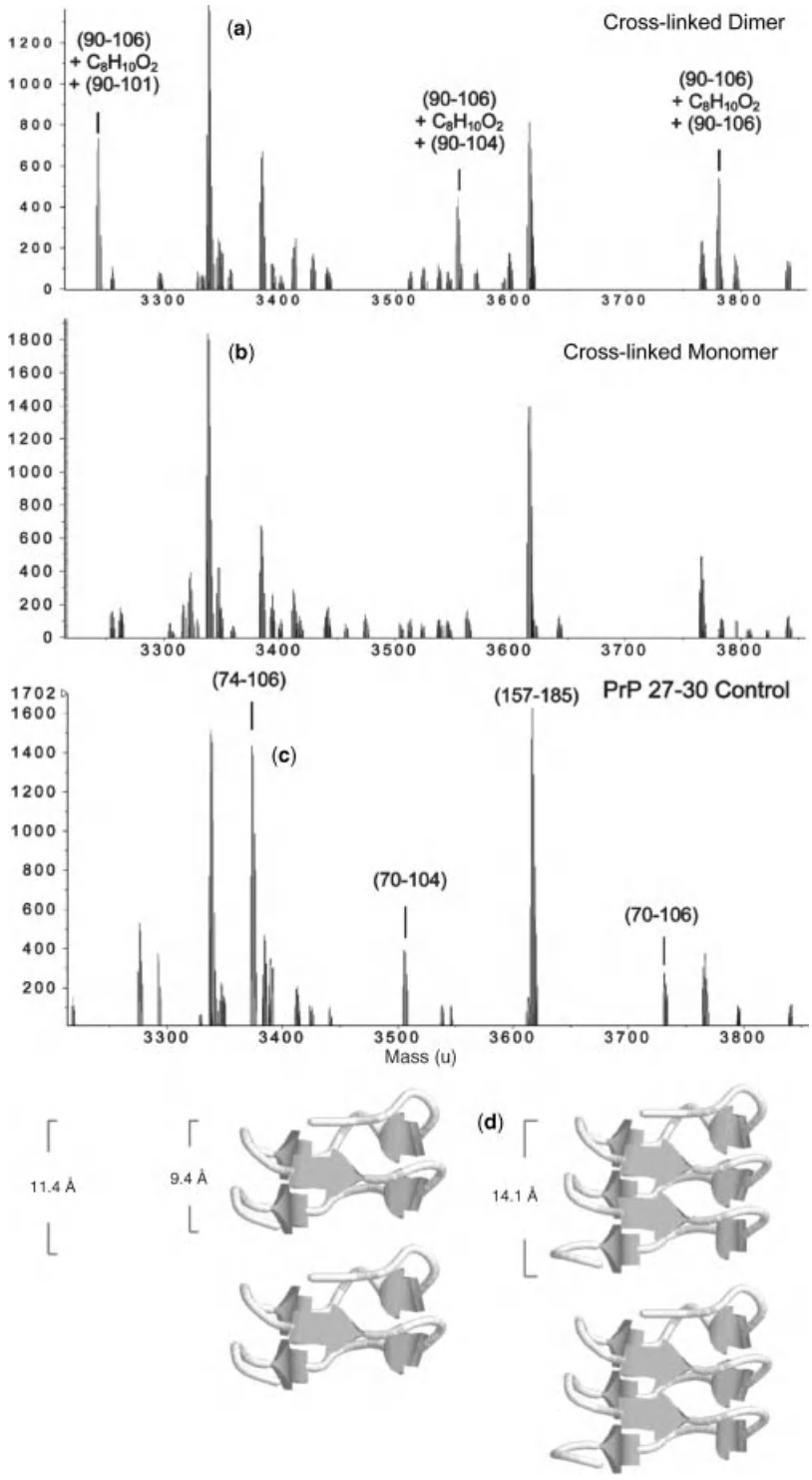


Figure 9.14. The MALDI TOF mass spectra of tryptic digests of the cross-linked dimer (*a*), monomer (*b*) and control deglycosylated PrP variant (*c*) obtained from bands excised from gels. The diagram in (*d*) shows compatibility of the detected Gly90–Gly90 cross-link with the model of the core of PrP based on vertical stacking of β-helices along the fiber axis. [Adapted with permission from (107). Copyright © 2005 American Chemical Society.]

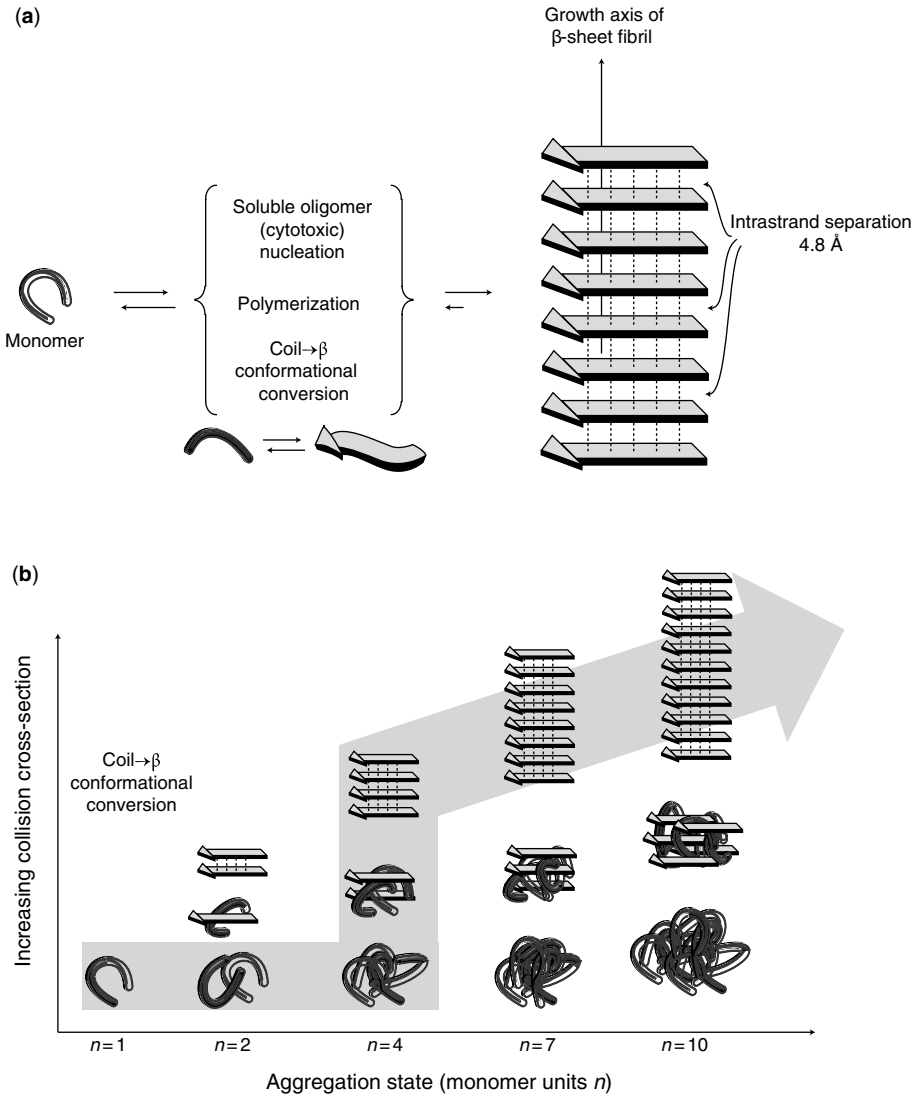


Figure 9.15. (*a*) Proposed processes during the soluble stage of peptide self-assembly in amyloid formation. Monomers self-assemble into mature, insoluble β-sheet fibrils through an intermediate phase of soluble oligomers. A pronounced transition of soluble oligomers from globular conformations into β-strand structures occurs during this phase. (*b*) Self-assembly starts at the folded monomer (left) and proceeds to soluble peptide assemblies of increasing mass (right). Soluble peptide oligomers with identical mass (i.e., number of monomer units *n*) can assume different conformations, such as globular (bottom row) or β-strand conformations (top row) with different collision cross-sections. Successively mass-extracting a specific aggregation state from the solution-phase distribution and subsequent determination of its collisional cross-section reveals the self-assembly pathway that occurred in solution (see large arrow). [Adapted from (110).]

enzymatic PTMs do not have the same level of precise control that makes the process of translation from genetic material so perfectly reproducible, and the end-products of polypeptide chain expression so homogeneous. Glycosylation is one particular form of enzymatic PTMs that is ubiquitous in higher organisms (also see Chapter 8). Structural heterogeneity exhibited by glycoproteins depends on the extent of glycosylation: Typically, even dealing with large proteins >100 kDa having <5% total carbohydrate content (by mass) is not problematic, while glycosylation levels exceeding 10% of the total protein mass poses serious problems.

Unfortunately, glycoproteins with carbohydrate content exceeding 10% of the total protein mass are not that uncommon in mammals; even higher extents of glycosylation are frequently seen in biopharmaceutical products. This poses a formidable challenge to MS based studies of higher order structure, dynamics, and interactions of such systems, which can be best appreciated in Figure 9.16.

This figure presents the results of studies carried out in our laboratory over 6 years ago with a candidate protein therapeutic designed to bind interleukin 1β in plasma (we refer to this protein as ILBP). In addition to a significant size (nearly 300 kDa), this protein has very high carbohydrate content (>20% by weight), which inevitably leads to a very convoluted appearance of the ESI mass spectrum (black trace in Fig. 9.16). In fact, the charge-state distribution of ILBP is totally unresolved, and it is impossible to determine the protein mass based on these measurements. Some features of the mass spectrum may suggest that a fraction of the protein fails to fold to a native conformation (leading to appearance of higher charge density ions at m/z 2000–4000), while folded molecules have a tendency to oligomerize (leading to appearance of ions at $m/z > 9000$). Furthermore, addition of the target ligand (IL1β) to the ILBP solution (gray trace in Fig. 9.16) results in subtle, but noticeable changes in the mass spectrum that could be interpreted as being the result of ILBP–IL1β binding. Nothing, however, could be said with certainty, and interpretation of the MS data is clearly speculative.

This example highlights the challenges that the researcher faces when working with highly heterogeneous systems. It is hardly surprising that such objects have not been among the most popular targets of MS studies aimed at understanding their conformational properties and interactions with physiological partners and therapeutic targets. Had the Editor insisted that a section on such systems be included in the previous edition of the book 7 years ago, it would probably contain a single sentence: This page was intentionally left blank. This situation is, however, changing very rapidly; a welcome change fueled in part by researchers' curiosity, and in large part by the growing demand in the biotechnology sector to develop robust analytical methods capable of dealing with highly heterogeneous systems. We already discussed in Section 4.1.2 a suite of methods designed to facilitate the analysis of native ESI MS data on glycoprotein assemblies. A combination of complexity reduction (by mass selecting a narrow fraction of glycoprotein ions) and gas-phase chemistry (charge reduction via electron transfer) was shown to be particularly useful in dealing with such systems (82). This same approach was also used successfully to study binding of highly glycosylated proteins to their physiological targets, (e.g., hemoglobin sequestration by haptoglobin) (82).

This approach can be used not only to study interactions within highly heterogeneous systems, which would be extremely helpful in trying to interpret the data shown in Fig. 9.16, but also to characterize their conformational properties. This is illustrated in Figures 9.17 and 9.18, which show the results of recent work carried out in our laboratory (114), with the goal of understanding properties of PEGylated interferon β1a (PEG–IFN). We have already looked at various aspects of interferon β1a structure and behavior under a variety of conditions in Chapters 5 and 7. PEGylation of this protein drug is intended to increase its

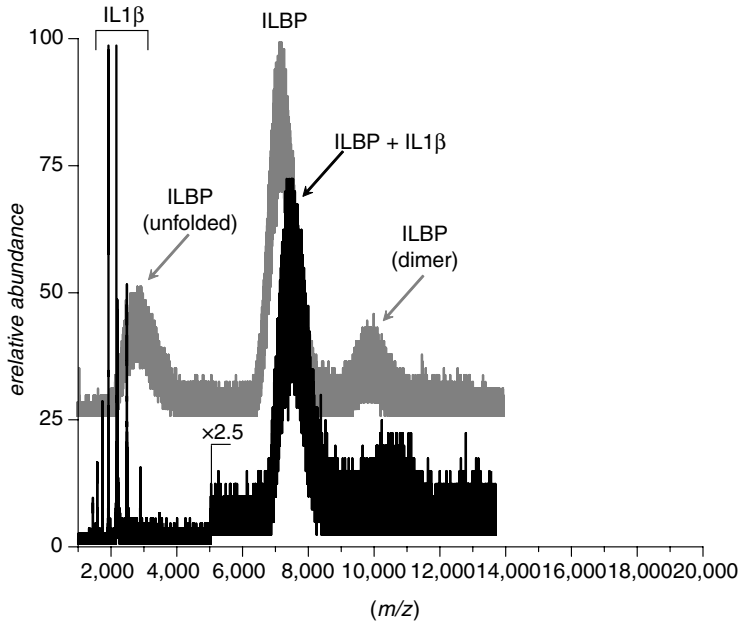


Figure 9.16. Electrospray ionization mass spectra of a highly glycosylated 270 kDa fusion protein ILBP (interleukin binding protein) with a total carbohydrate content exceeding 20% by weight (gray trace) and ILBP mixture with excess of interleukin 1β (IL1β), its intended target (black trace). Both spectra were acquired under near-native conditions. [Data courtesy of Dr. Joshua K. Hoerner, presently at Schering-Plough).]

circulation lifetime without altering its conformation which, as seen in Chapter 7, is absolutely critical for binding to its physiological receptors. While conformational integrity of interferon β1a covalently modified with small functionalities is easy to evaluate by monitoring ionic charge-state distributions in ESI MS (see, e.g., mass spectra of the intact and NEM-modified proteins in Fig. 5.2), ESI MS data of PEG–IFN are not so easy to interpret. Despite the relatively modest molecular weight of the conjugate (∼ 45 kDa for the mono-PEGylated product), the polydispersity exhibited by the 20-kDa PEG chain gives the ESI mass spectrum of PEG-IFN a very convoluted appearance (gray trace in Fig. 9.17). Apparent overlapping of ionic peaks representing different charge states, and different lengths of PEG chains, results in a totally unresolved spectrum; however, selecting precursor ions within a narrow *m/z* window followed by partial reduction of their charge induced by electron transfer (black trace and inset in Fig. 9.17) allows the average mass of PEG–IFN to be measured with great confidence.

Once the average protein mass is known, charge distributions can be plotted for PEG–IFN ions in ESI mass spectra acquired under near-native and denaturing conditions, which are essentially inverted mass spectra (Fig. 9.18). The difference between these distributions and the ones we dealt with in Chapter 5 (e.g., see Fig. 5.3) is their non-discrete character. However, these continuous distributions can be processed in much the same way as discussed in Section 5.1, when presenting the deconvolution approach to analyzing conformational dynamics of small homogeneous proteins (115,116). In fact, only a single Gaussian curve is needed to fit the profile of PEG–IFN ions in the mass spectrum acquired under native conditions (black trace and light gray curve in Fig. 9.18*b*). As Houde et al. (117) demonstrated recently using HDX MS, the protein core in PEG–IFN remains intact; therefore, the Gaussian curve that fits the "native" PEG–IFN charge-state distribution can be assigned as a basis function representing a stable protein core and a floppy polymer chain. If we use this curve as a fixed basis function, only one additional curve was needed to generate a good fit of experimental data for the denatured PEG–IFN charge-state distribution (darker gray trace in Fig. 9.18*b*). This curve represents the contribution of less structured protein–polymer conjugates to the overall ionic signal. Interestingly, deconvolution of the charge-state distribution of intact interferon β1a under the same conditions also requires only two basis functions, representing a highly structured native conformation and a less structured state. The presence of PEG chains does not appear to introduce any additional states under either near-native or denaturing conditions, suggesting a fairly benign role played by the polymer chain in modulating protein structure (114). What is perhaps most remarkable, is the wealth of information that can be deduced from the seemingly uninterpretable MS data shown in Fig. 9.18!

This section would be incomplete without mentioning the great potential of ion mobility MS as a means of dealing with highly heterogeneous systems. While the majority of current applications of this technique utilize its unique capability to provide information on the geometry of macromolecular ions in the gas phase, the potential utility of

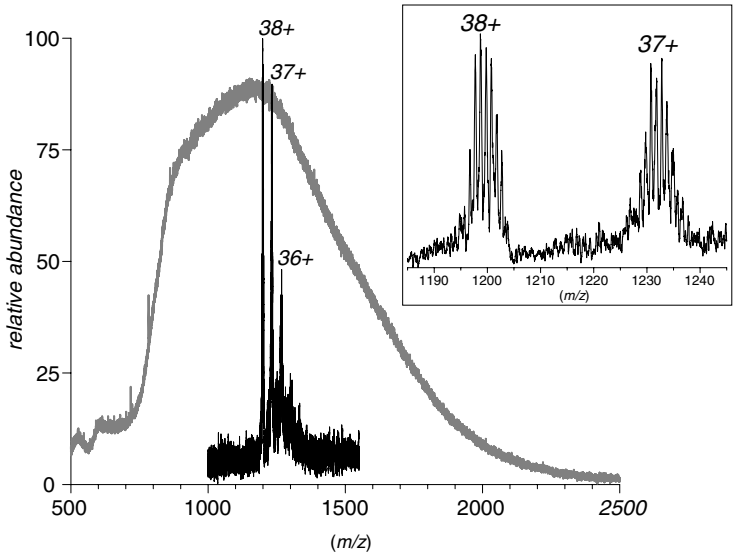


Figure 9.17. The ESI MS of polylethylene glycol (PEG)–IFN acquired under denaturing conditions (gray trace) and the result of charge reduction within a narrow selection of conjugate ions induced by electron transfer (black trace). The inset shows an expanded view of two products of the charge reduction process, which allow the average mass of the protein–polymer conjugate to be assigned as 45.6 kDa. [Data courtesy of Dr. Rinat R. Abzalimov (UMass-Amherst).]

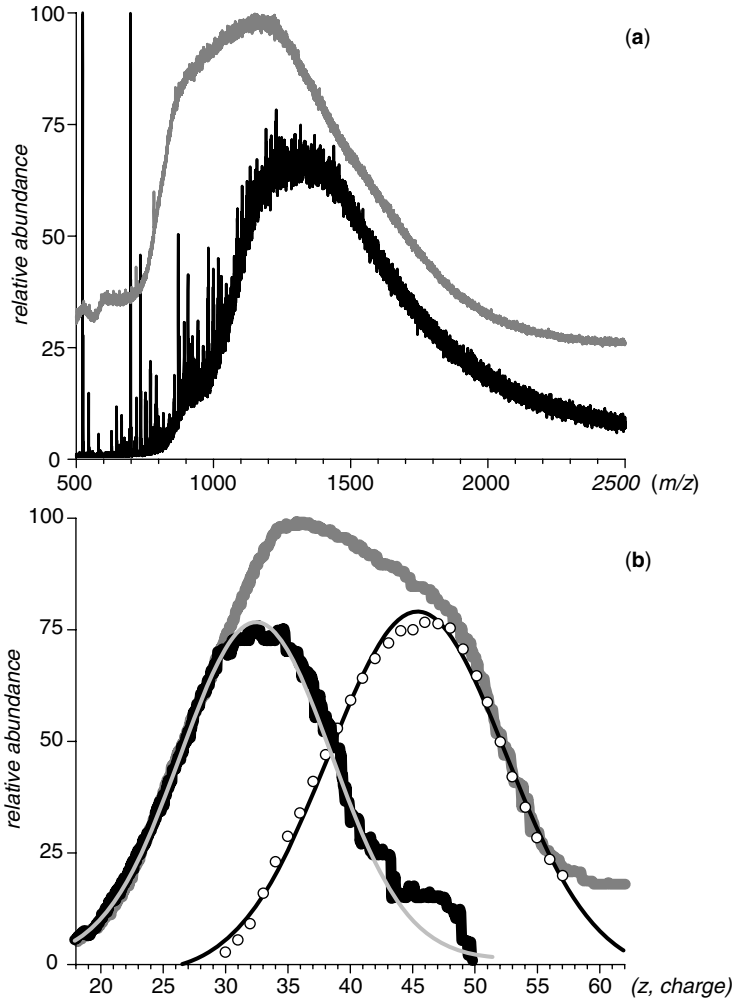


Figure 9.18. (*a*) The ESI mass spectra of PEG–IFN acquired under near-native (black trace) and denaturing conditions (gray trace). (*b*) Continuous charge state distributions of PEG–IFN extracted from the ESI MS data (*a*) using the mass information provided by charge reduction experiments shown in Figure 9.17. The charge distribution of "native" PEG–IFN (black trace) fits to a single Gaussian curve (light gray). Using this curve as a fixed basis function allows the charge distribution of "denatured" PEG–IFN (gray trace) to be fitted using only one additional Gaussian curve, which is assigned as a basis function representing less structured PEG–IFN molecules. [Data courtesy of Dr. Rinat R. Abzalimov (UMass-Amherst).]

IM–MS as a means of reducing complexity of heterogeneous systems by providing an additional separation stage prior to MS detection is frequently overlooked. The ability of this technique to separate various isoforms of both carbohydrates (118) and synthetic polymers (119) is well documented, and in fact has already been used to facilitate the analysis of the covalent structure of large glycoproteins (120) and protein–polymer conjugates (121). We anticipate that continuous improvements of this technique will make it an indispensable tool in the analysis of various aspects of behavior of glycoproteins, protein–polymer conjugates, and perhaps even more complex systems, (e.g., functionalized nanoparticles).

9.4. HOW LARGE IS "TOO LARGE"? MASS SPECTROMETRY IN CHARACTERIZATION OF ORDERED MACROMOLECULAR ASSEMBLIES

Although the examples of catastrophic protein aggregation considered in Section 9.2 lead to pathological conditions with grave physiological manifestations, controlled large-scale protein association processes are crucial for normal cell functioning. Understanding how the components of sophisticated biological machinery work together is critical not only for satisfying our curiosity by understanding molecular mechanisms of various biological phenomena, but also for a variety of practical applications, such as

designing new efficient therapeutic strategies to treat a range of pathological conditions. Above and beyond protein–protein interactions, a variety of other biopolymers participate in forming critical nodes within the sophisticated interactomes. In addition to protein–oligonucleotide binding events playing obviously important roles at the terminal points in such networks (e.g., gene expression), a range of other interactions provide important mechanisms to transmit, suppress or modify the signals both inside and outside the cell, such as encounters involving noncoding ribonucleic acid (RNA) (122,123) and glycosaminoglycans (124). Large protein assemblies play a crucial role in cellular function. In fact, each cellular protein emerges from a large assembly upon its birth (ribosome) and enters another large assembly at the end of its "life" (proteasome), while interacting (or becoming a part of) a number of other assemblies in between these two termini. In this section, instead of following the entire cradle-to-grave pathway we will consider how MS can help us to understand the functions of biological machinery representing a few points on this pathway, by providing a wealth of information on relevant macromolecular assemblies.

9.4.1. Proteasomes

The proteasome is a large multiunit proteolytic machine that is a key element of regulated degradation of various cellular proteins, a process needed for maintaining cellular homeostasis. We already briefly mentioned (Section 4.1.5) the pioneering work by Loo et al. (125), who used ESI MS with an ion mobility GEMMA device to characterize the structure of a noncovalent 28-subunit complex representing the 20S proteasome. The ESI MS measurements confirmed both the stoichiometry and overall organization of this gigantic complex by providing mass information on a single α_7 ring (192 kDa) and the intact $\alpha_7\beta_7\beta_7\alpha_7$ complex (690 kDa). Importantly, the authors of this study were able to see not only binding of a reversible inhibitor to the proteasome, but also sequestration of multiple model substrates by the inhibited proteasome (125).

Although the proteasome structure has been elucidated by X-ray crystallography, it does not reveal the mechanisms underlying substrate binding and translocation of polypeptide chains into the proteasome interior; even the number of substrate molecules that are sequestered within the internal cavities at any given time is unknown. To fill this gap, Robinson and co-workers (126) applied both electron microscopy and MS/MS to obtain precise information regarding the stoichiometry and localization of substrate proteins within the proteasome chambers. The dissociation pattern of this macromolecular assembly in MS/MS experiments (Fig. 9.19) allowed the highest number of each substrate molecule bound within the cavities to be established. Remarkably, > 95% of all proteasome molecules contained the maximum number of partially folded substrates. A detailed analysis of the complex fragmentation pathways allowed partitioning of different substrate proteins between the central proteolytic chamber and the antechambers to be established. An intriguing conclusion of this work was that in addition to protein degradation, 20S proteasomes can also act as a substrate storage facility prior to their degradation, specifically in cases where translocation rates are slower than proteolysis (126).

9.4.2. Ribosomes

Ribosomes are cellular components that execute the final step in the gene expression pathway by catalyzing messenger RNA (*m*RNA)-directed protein synthesis (127,128). Elucidation of the mechanism of protein synthesis has been a central problem in structural biology for several decades. While the crystal structures of bacterial ribosomes have been obtained (see (129) for a historical account), many questions still remain as far as the functioning of this formidable cellular apparatus in both prokaryotic and eukaryotic organisms (130,131). The general features of ribosomal structure and functional mechanism are conserved in all cellular forms of life. Among all ribosomes, the prokaryotic ones are the smallest (*only* 2.5 MDa) and are the best studied. They are on the order of 250 Å and are composed of two subunits, 30S and 50S (the latter about twice the size of the former: Subunits are denoted by their apparent sedimentation coefficient expressed in svedberg units). The 30S subunit consists of a single RNA molecule (16S *r*RNA), which is $\sim$ 1500 nucleotides in length, and single copies of each of $\sim$ 20 different protein molecules (Fig. 9.20). The 50S subunit consists of a large RNA molecule (23S *r*RNA, which contains $\sim$ 2900 nucleotides), a smaller molecule of RNA (5S *r*RNA, $\sim$ 120 nucleotides), and $\sim$ 30 different proteins (Fig. 9.20). Although their architectural organization is very similar, eukaryotic ribosomes are larger compared to those of prokaryotic organisms. Generally, they contain more components, and each component's size is larger than that of its prokaryotic analogue; intact prokaryotic ribosome sediments at 70S.

The initial attempt to use MS as a structural tool in the study of ribosomes was made by Robinson and coworkers (134), who used nano-ESI to successfully transfer an intact *E. coli* ribosome from solution to the gas phase with consequent mass analysis of its constituents. Although gas-phase dissociation of the intact ribosome could not be avoided in this study, the partial disassembly processes were remarkably selective, showing strong correlation with the predicted features of ribosomal protein–protein and protein–RNA interactions. In addition, the observation of distinct peaks corresponding to individual protein components in the ESI mass spectra provided an opportunity to probe the conformational and dynamic properties of these proteins.

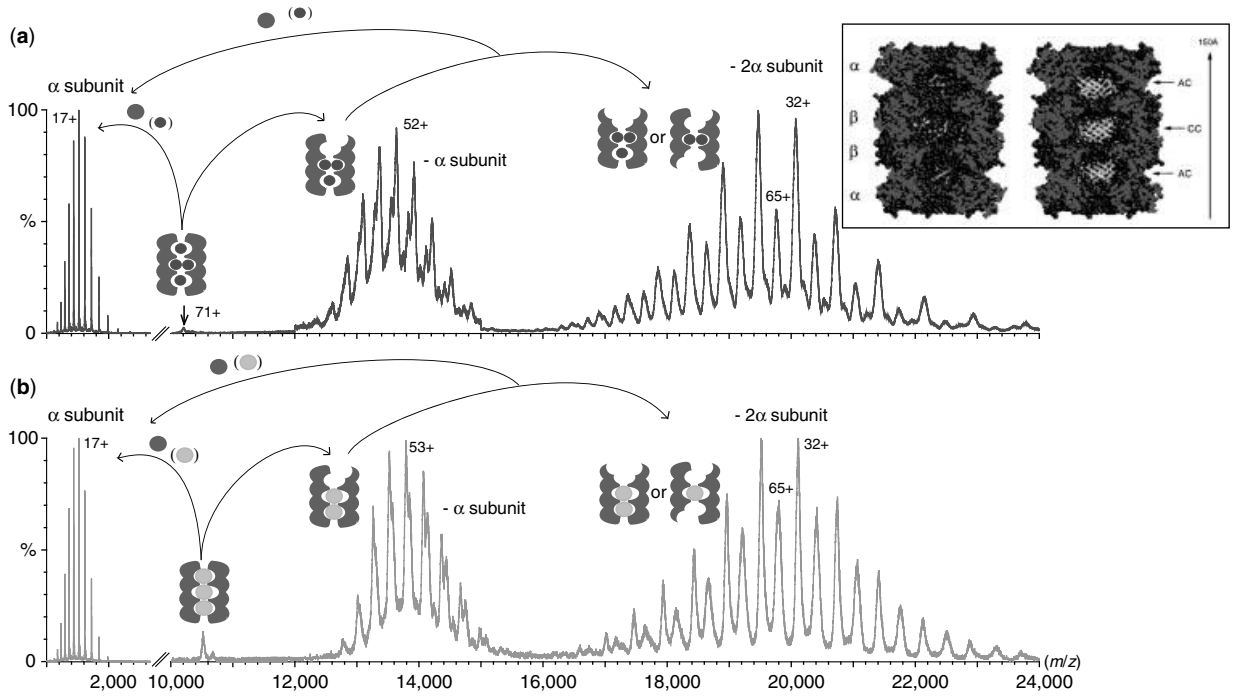


Figure 9.19. Schematic representation of the dissociation pattern of the host–guest complexes during MS/MS. Fragmentation spectrum of the proteasome–cytochrome *c* (Cyt *c*) host–guest complex is shown in (*a*) and with green fluorescent protein (GFP) in (*b*). Initially four Cyt *c* or three GFP molecules bind within the proteasome. During the first dissociation step, the stripping of one α-subunit exposes the substrate molecule within the antechamber and triggers its dissociation. In the second dissociation step, an additional α-subunit is lost, either from the same ring or from the opposing one. In the case that the α-subunit dissociates from the opposing ring, another substrate molecule is exposed and released. The stripping of an additional α-subunit from the same ring will not trigger further dissociation of substrate molecules. Inset: space-filling model of the proteasome host–guest complexes. The substrates Cyt *c* (*a*) and GFP (*b*) are placed within the two antechambers and the catalytic cavity to illustrate their fitting. For clarity, 50% of the proteasome complex was removed to show the inner cavities. The substrates are shown in their folded conformation, although the experimental data suggest that the substrates are less compact than the native state. [Reproduced with permission from (126). Copyright © 2006 American Society for Biochemistry and Molecular Biology.]

Many of these proteins are known to be at least somewhat unfolded when removed from the ribosome, and very little is known about their dynamics within the context of the intact particle. The HDX measurements carried out on 50S subunits indicated that many proteins are highly protected. For example, a very high degree of protection was exhibited by L11 protein, which is consistent with a highly folded conformation assumed by this protein within the ribosome. In fact, the extent of L11 protection was so high for a protein of its size (14 kDa), that it matched the protection levels characteristic of similarly sized proteins in crystals, rather than in solution (134). Intriguingly, earlier NMR measurements indicated that a C-terminal fragment of protein L11 (hosting the RNA binding domain) is significantly destabilized in the absence of *r*RNA (135). The remarkable difference between the two states of L11 suggests that tight packing within the intact ribosome restrains the structural fluctuations of otherwise flexible segments of the protein (134). This initial study demonstrated the power of ESI MS as a tool to probe not only the topological features of ribosomal assemblies, but also the dynamic behavior of its constituents within the context of the intact ribosomal particle.

The absence of intact ions of the ribosome or either of its subunits in the ESI spectra of *E. coli* ribosome (134) was initially interpreted by many as a proof that such large complexes cannot survive the transition from solution to the gas phase. However, our earlier discussion in Section 4.1.4 of the relationship between a protein's size and the number of charges it accumulates in the gas phase offers a different explanation. The number of charges expected to be accumulated on an $\sim$ 250 Å particle is not significant enough to produce ions within the *m/z* range used in the initial study (134). Spectacular advances made by both ESI and

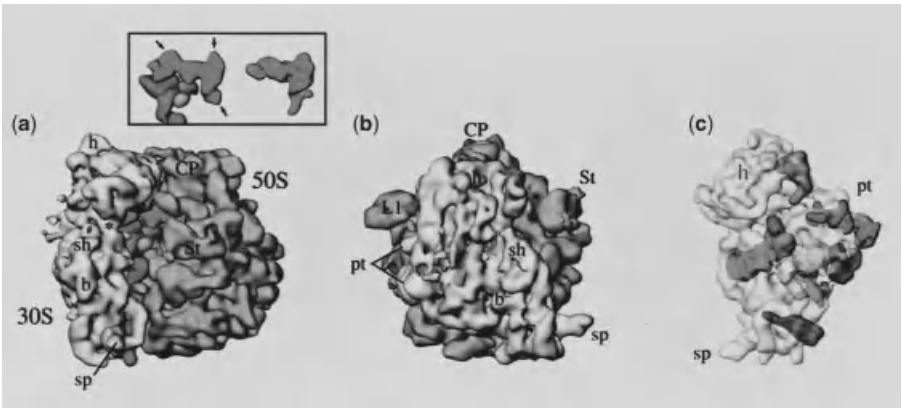


Figure 9.20. *Escherichia coli* initiation complex, with transfer–RNA fMet–tRNA bound to the ribosome in the presence of mRNA with an AUG codon in its middle. (*a*, *b*) Two views of the 70S ribosome. Inset in (*a*) shows on the left the experimental mass attributed to tRNA in contact with the decoding center, and on the right, atomic structure of initiator tRNA, filtered to 10 Å resolution. Arrows point to contacts between tRNA and ribosome. (*c*) The 30S portion of the map, with fitted X-ray structures of proteins. [Adapted with permission from (132) as presented in (133). Copyright © 2003, John Wiley & Sons Ltd.]

time-of-flight (TOF) technologies in the late 1990s led to a dramatic expansion of the *m/z* range within which protein ions could be detected. This allowed the Robinson group (136) to observe ions corresponding to the intact 30S subunit of an *E. coli* ribosome. Remarkably, the mass deduced from the ESI MS data was only 5.5 kDa higher than that calculated based on the sequences of the 21 proteins and the 16S RNA molecule (846.7 kDa). Mass spectra of an intact subunit 50S and intact ribosome (70S) were acquired as well; however, larger macromolecular ion size and the need to use 10 mM magnesium acetate to stabilize this assembly led to significant overlap of ionic peaks in mass spectra, which prevented charge assignment from being done in a straightforward manner (136).

Once the feasibility of intact ribosome detection was demonstrated, Robinson and co-workers (137) focused their attention on investigating various factors that govern the loss of proteins from ribosomes in the mass spectrometer. Such information can be used to probe dynamic events in the ribosome induced by binding of various cofactors (elongation factor G, whose binding triggers translocation, etc.). Mass spectrometry was used to identify proteins that are released in the gas phase from *E. coli* ribosomes in response to varying solution conditions and cofactor binding. Protein components that dissociate most readily from the ribosome were identified and such dissociation patterns were analyzed against the available structural information for ribosomes (the structural details of the protein–RNA interactions in the *E. coli* ribosome were inferred from the high-resolution atomic structures of a ribosomal subunit 50S from *Haloarcula marismortui* and intact 70S from *Thermus thermophilus*, an approach justified by the high degree of structural conservation across bacterial ribosomes). The results of such analysis strongly suggested that a high interaction surface area of the protein with RNA (measured as ΔSASA, change in solvent accessible surface area) was the major force in preventing dissociation. Gas-phase dissociation was enhanced by loosening the structure of the intact ribosome in solution as a result of pH changes, metal ion content variation, and so on. This relationship between the protein–RNA ΔSASA and the propensity for gas phase dissociation was used to interpret the release of the protein components from the ribosome, which was complexed with the elongation factor G (EFG). Two inhibitors of elongation were used in these experiments, fusidic acid and thiostrepton, freezing the ribosome in complexes analogous to post- and pre-translocation states, respectively. While the presence of fusidic acid in solution did not result in significant alteration of the ribosome dissociation pattern, thiostrepton induced major perturbations in protein–RNA interactions, as suggested by dramatic changes in the ribosomal dissociation pathways. The results of these experiments clearly demonstrate that the two ribosome–EFG complexes are significantly different and are indicative of conformational changes occurring during translocation (137).

This approach of utilizing MS/MS to aid understanding of the spatial organization of ribosome components was later extended to eukaryotic systems (138,139). An interested reader is referred to an excellent minireview by Gordiyenko and Robinson (140) summarizing the studies of ribosomal structure using direct ESI MS measurements. In addition to direct ESI MS and MS/MS, ribosomal structure has also been probed recently using other techniques discussed in Chapter 4, such as chemical labeling (141), cross-linking (142), and HDX MS (143,144). For example, Reilly and co-workers (141) demonstrated that the extent of

labeling of ribosomal components is strongly dependent on the solvent accessibility of the target residues (141). Proteins of the ribosome stalk, which are known to be largely solvent accessible, were labeled extensively, while the proteins shielded from the solvent in their subunits had much lower extents of derivatization. Moving from structure alone to conformational dynamics, Yamamoto et al. (143) used HDX MS to measure deuterium incorporation in 55 proteins constituting the 30S and 50S ribosomal subunits, to elucidate the regions and the magnitude of structural fluctuations in the 70S ribosome from *E. coli*. In a more recent study by the same group, the influence of Mg^{2+} on the flexibility of various components of the 70S ribosome was investigated with HDX MS. This metal cation is needed not only to maintain the sophisticated architecture of the 70S ribosome, but also to enable protein synthesis by providing a delicate balance between conformational stability and flexibility. The HDX MS measurements carried out in this work provided information necessary to appreciate the complex relationship between ribosomal structure, flexibility, and function.

Finally, an intriguing recent study by Ott and coworkers (142) used chemical cross-linking as a means of visualizing the workings of a mitochondrial ribosome from yeast. The study focused on the events in the vicinity of the polypeptide tunnel exit, where nascent chains are exposed to an aqueous environment for the first time. This work revealed a complex network of interacting macromolecules, including proteins and protein domains specific to mitochondrial ribosomes. This network was shown to include a membrane-bound ribosome receptor of the inner membrane, as well as several ribosomal proteins exclusively found in mitochondria.

9.4.3. Molecular Chaperones

While ribosomes and proteasomes considered in the preceding sections are large macromolecular assemblies representing the two terminal points in the life of almost every cellular protein, there are many other mega complexes with which proteins interact during their life cycle. One such type of large protein assembly that has received much attention is represented by a class of helper proteins, known as molecular chaperones. These proteins are thought to mediate the process of folding in the cell, in some cases by maintaining the chains in an extended conformation in order to prevent premature (mis)folding, and in other cases to provide an environment conducive to efficient renaturation without aggregation. A number of different classes of chaperones are known, first identified as *heat-shock proteins* (Hsp's) due to their increased expression under stress conditions, such as elevated temperature, where newly expressed proteins may need more assistance in correct and efficient folding. One of the most interesting classes of chaperone proteins, at least in terms of studies by MS, is the Hsp60 proteins, specifically GroEL in bacterial cells. This protein, as its name suggests, is comprised of 60 kDa monomers, but these are assembled into a toroidal structure, the active form of the protein being a macromolecular assembly of two stacked heptameric rings (Fig. 9.21), with an overall

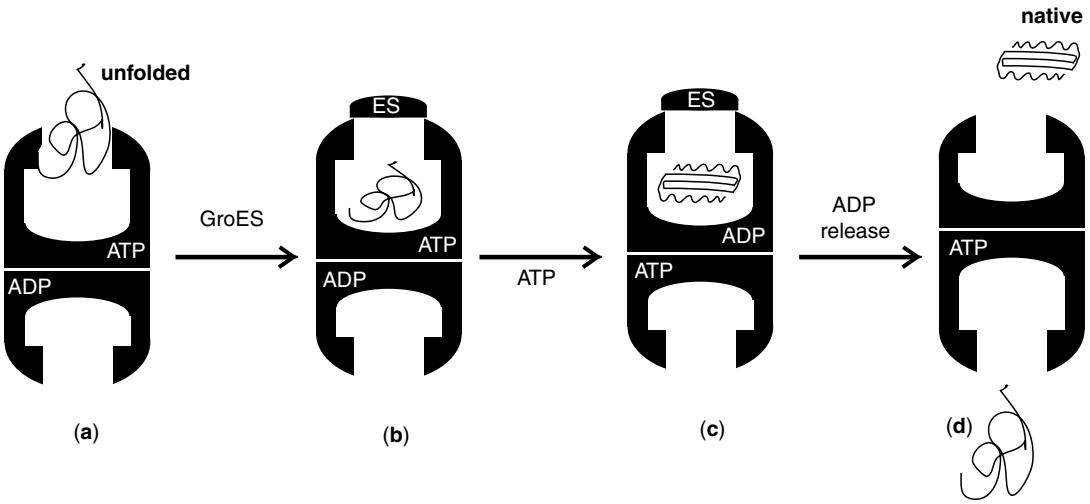


Figure 9.21. Schematic of the "cis" GroEL–GroES reaction cycle (145). (*a*) Upon binding of substrate peptide followed by (*b*) capping of the cavity by GroES, (*c*) *cis*-ATP hydrolysis triggers a conformational change within the cavity shifting the interior surface from hydrophobic to hydrophilic in nature. This promotes protein refolding. (*d*) The ATP binding to the opposite side of the GroEL machine triggers release of GroES and folded substrate, and prepares the opposite ring for the next peptide binding event.

molecular weight of $\sim$ 800 kDa (we have already seen a mass spectrum of such an assembly in Fig. 4.4). The function of this structure is to provide a hydrophobic cavity within which partially unfolded proteins can be sequestered, unfolded, and then allowed to refold in an environment where aggregation is disfavored. This complex process requires large conformational changes in the GroEL machine and is known to be assisted by the cochaperone protein GroES, a single ring heptamer of Hsp10 proteins, with energy provided by hydrolysis of Mg–ATP.

Mass spectrometry has been extremely helpful in understanding the molecular mechanisms that govern the work of these cellular "correctional facilities". For example, it provided critical information that allowed the conformational state of an entire protein bound within the GroEL cavity to be determined. If the chaperone-bound polypeptide is allowed to undergo HDX reactions, then this can give some valuable information as to whether the bound substrate is folded, completely unfolded, or adopts some conformational ensemble that is intermediate between the two. Upon desolvation into the gas phase, the proteins will dissociate from one another, but the hydrogen-exchange information will be retained. Exchange experiments performed on a three-disulfide derivative of α-lactalbumin showed that binding of this protein to GroEL caused the protein to adopt a conformation that was at least somewhat protective against exchange (146), reminiscent of the molten globule state known to be an equilibrium unfolding intermediate for this protein (147). In these experiments, three complementary pieces of information could be obtained. From the mass of the substrate protein the degree of protection against exchange could be measured, indicative of the conformation of the bound peptide. Additionally analysis of the charge-state distribution of the substrate protein suggested a partially folded species. Finally, some idea of the conformational heterogeneity was obtained by measuring the width of the exchange peaks. If the partially folded state adopted a wide range of different conformations, then this would manifest itself as a very broad distribution of masses, whereas what was actually observed was a quite narrow mass peak, indicating that the GroEL bound species, although only partially folded, retained persistent secondary structure (148).

Similar studies by NMR and MS using the protein dihydrofolate reductase (DHFR) suggested that, in some cases, the conformational states bound to GroEL may even be quite highly native-like (149,150). These studies could distinguish no difference between protection of the substrate protein as it bound initially and in subsequent binding–release cycles. This suggests the possibility that GroEL assists refolding by forcing subtle structural changes of the partially folded substrate protein during iterative binding and release, as opposed to a wholesale unfolding of the substrate upon binding. Using hen lysozyme as a substrate, no change in folding mechanism was observed, although the overall rate of folding for this protein was accelerated in the presence of chaperone (151).

Plückthun and co-workers (152) carried out extensive work to elucidate the effect of GroEL on the folding and unfolding pathway of β-lactamase. They observed very high protection against HDX of β-lactamase in the GroEL bound form. Additionally, the measured unfolding rate from the bound state was measured to be identical to that of the native state (unfolding of the equilibrium molten globule state occurs too rapidly to be measurable). The authors concluded that different proteins are recognized by GroEL in very different states, ranging from totally unfolded to native-like, depending on the accessibility of exposed hydrophobic patches. Reversible binding of native-like states may be an important property of GroEL in protection of proteins against aggregation under heat-shock conditions. Additional studies by this group used limited proteolysis at elevated temperature to identify the binding sites on β-lactamase that are recognized by GroEL (153), and also used different denaturing conditions to identify different conformations of the protein that could be bound to the chaperone (154).

Further insight into the action of GroEL was gained from a kinetic study of the folding of malate dehydrogenase (MDH) in the cavity of a single ring variant of GroEL (SR1) in the presence of the cochaperone GroES (155). Malate dehydrogenase does not fold efficiently in the absence of chaperones, and apparently requires several rounds of binding and release from the GroEL–GroES cavity for complete recovery of native structure. The authors investigated the conformation of non-native MDH bound to SR1 alone using HDX MS, which showed a core-like structure that was significantly protected against exchange even in this non-native state. Some loss of protection was observed upon addition of GroES and ATP, but this could be attributed to the breaking of hydrogen bonds between substrate and the walls of the chaperone cavity rather than a forced unfolding reaction. This is consistent with the notion that the binding of cochaperone induces a conformational change in the cavity that releases the unfolded polypeptide into a hydrophilic environment, which is permissive for protein refolding. It does not support the alternative hypothesis of a mechanical unfolding process prior to release. A global view of chaperone activity can be provided by direct ESI MS measurements in suitable solvent systems. Heck and co-workers (156) used this approach to analyze macromolecular complexes involved in the chaperonin-assisted refolding of gp23, a major capsid protein of bacteriophage T4. It was found that GroEL can bind up to two unfolded gp23 substrate molecules; however, when GroEL was in complex with the cochaperonin gp31, it was capable of binding exclusively one gp23 molecule. This work also demonstrated that folding and assembly of gp23 into 336 kDa hexamers by GroEL–gp31 can be monitored by ESI MS.

9.5. COMPLEXITY OF MACROMOLECULAR INTERACTIONS *IN VIVO* AND EMERGING MASS SPECTROMETRY BASED METHODS TO PROBE STRUCTURE AND DYNAMICS OF BIOMOLECULES IN THEIR NATIVE ENVIRONMENT

9.5.1. Macromolecular Crowding Effect

The examples of using various MS-based approaches to study the mechanisms of sophisticated biochemical machinery presented in Section 9.4 highlighted the power of this technique to provide information critical for understanding the molecular basis of complex biological events. However, even though studying protein behavior under carefully defined conditions in the test tube is a very convenient method, there was always a significant amount of controversy as to its physiological relevance. In the intracellular environment that proteins are exposed to as they are translated and released from the ribosome, they are by no means isolated species allowed to interact with a carefully selected set of partners at low concentration. Instead they are released into an extremely crowded intracellular environment containing a variety of other proteins, nucleotides, ions, and at a significant osmotic pressure, an environment where the collective behavior of all players was eloquently named "molecular sociology of the cell" (157). Thus, there is significant interest in determining not only the mechanistic details of protein folding and interactions *in vitro*, but also the similarities and differences *in vivo*. One of the major differences is that the total concentration of macromolecules in a typical *in vivo* setting is very high, reaching levels of 50–400 mg/mL in cellular interiors (158). Since all species occupy a significant proportion of the total volume of the medium (regardless of how insignificant the contribution of any given macromolecule may be), such media are referred to as "crowded" rather than "concentrated" (158–161). Taken together, the macromolecules occupy a significant fraction of the total volume, making it physically unavailable to other macromolecules (Fig. 9.22).

In such a crowded environment, nonspecific interactions between macrosolutes may exert very significant influence on their properties (the *macromolecular crowding effect*). Hall and Minton (161) point out that the concept of "nonspecific interactions" is widely misunderstood (162). These interactions are often considered as artifacts that interfere with the acquisition of meaningful data and should be reduced or totally eliminated (e.g., by extrapolating the measurement results to zero macromolecular concentration). Although such procedures may be useful for understanding the behavior of isolated macromolecules, they do not necessarily provide results that are more meaningful in a

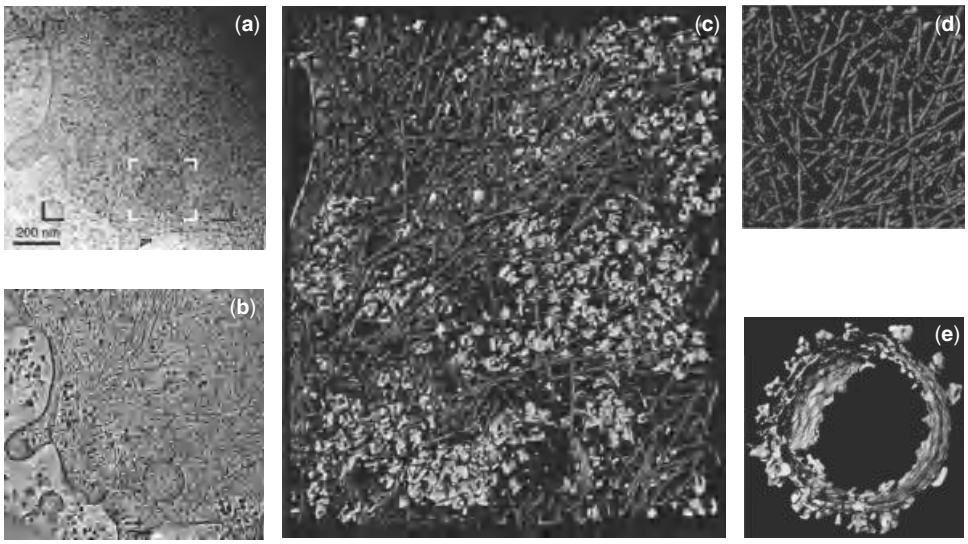


Figure 9.22. Macromolecular crowding of eukaryotic cytoplasm visualized by cryoelectron tomography. Peripheral region of a vitrified *Dictyostelium* imaged with conventional transmission electron microscopy, cell thickness 200–350 nm (*a*) and a representative tomographic slice of 60 nm thickness, revealing a cortical actin network (*b*). Visualization (surface rendering) of actin network, membranes, and cytoplasmic macromolecular complexes (*c*) in a $815 \times 870 \times 97$ nm volume corresponding to the area in panel *a* framed in black. An image of an actin layer (*d*) taken from the upper left portion of (*c*) reveals a crowded network of branched and cross-linked filaments. Surface rendering of a ribosome-decorated portion of endoplasmic reticulum segmented into membrane and ribosomes (*e*). The densely packed structures in the lumen are omitted for clarity. [Adapted from (162).]

biological context (158,159,163). Nonspecific interactions caused by crowding are unavoidable *in vivo*, as the crowding itself is an intrinsic property of most or all physiological fluid media. Therefore, many important aspects of macromolecular behavior in such media cannot be understood without taking the nonspecific interactions into account. The main effect of macromolecular crowding on biochemical equilibria is to favor the association of macromolecules. Crowding usually favors aggregation, particularly for slow-folding chains. The effect is greater for small polypeptides, since larger proteins have slower diffusion and, therefore, reduced encounter rates (163).

The effect of macromolecular crowding on protein behavior can be modeled by using high concentrations of macromolecules that mimic those found in living cells and do not directly interact with the macromolecules being investigated, except via steric repulsion (163). An ideal crowding agent should be fairly large ($\sim$ 50–200 kDa), highly water soluble and not prone to self-aggregation; globular (rather than extended) shapes may prevent the solutions from becoming too viscous. Commonly used synthetic crowding agents include dextrans, Ficolls (synthetic polymers of sucrose), PEG, and poly(vinyl alcohol) (PVA). Unfortunately, utilization of any of these crowding agents in ESI MS studies at relevant concentrations seems to be problematic due to the protein ion signal suppression (as discussed in this Section 9.1.2). Another serious problem is caused by the intrinsic heterogeneity of synthetic polymers giving rise to a wide distribution of ion peaks in the ESI mass spectra, which would almost certainly interfere with the protein ion signal. This problem may be eliminated by using inert ubiquitous proteins (e.g., lysozyme, serum albumin, or hemoglobin) as crowding agents. Nano-ESI may become an indispensable tool in such studies due to its increased tolerance to high concentrations of proteins in solution (164).

9.5.2. Macromolecular Properties *In Vitro* and *In Vivo*

Another important aspect of macromolecular behavior *in vivo* is the extreme complexity of the interactions due to the very high number of participants in such interactions. Throughout most of this book, our attention has been focused on the studies that use the *reductionist approach* to the problem. Each biological process is considered as a set of well-defined steps, or interactions in which the number of players is reduced to a minimum. The advantages of this approach are obvious, since it allows a huge many-body problem to be substituted with the analysis of properties of individual key biomolecules and their interactions with a limited number of key partners. The implicit assumption here is that if the function and properties of each individual component of a system are characterized, then the functioning of the entire system can be understood by a simple superposition. As pointed out by Westerhoff and coworkers (165), this "*principle is true in theory, but in practice is often false*" because of a widespread nonlinearity in biological systems at every level of organization. Nevertheless, until very recently the reductionist approach remained a commanding methodological dogma not only in molecular biophysics, but also in the entire field of structural biology.

The situation is now changing dramatically. Recent years have brought the realization of the limitations of the reductionist methodology in the life sciences (165–167). The completion of the entire genome sequencing of several organisms highlights the enormous repertoire of biomolecules that make up living cells. However, simple cataloging of the DNA and protein inventory alone does not automatically generate understanding of how a living cell functions. At the same time, it is unlikely that such knowledge will be provided by characterization of the individual properties of even a large number of individual biomolecules. It appears that there is need for a synthesis of the two approaches that would take our understanding of the complex networks of cellular processes to a new level (168–171). One approach to dealing with the complexity of real living systems that enjoyed great popularity in the past decade, is functional proteomics. Since the limitations of this book do not allow us to provide a detailed discussion of various methodologies that are currently in use in proteomics, an interested reader is referred to several extensive review articles (172–178) on this subject.

Apart from the proteomic approaches that give a bird's eye view of biomolecular interactions in living systems, no examples of using MS based strategies to study structure and dynamics of proteins in cells and beyond were available when the previous edition of this book was published. However, the situation is changing very rapidly, and a number of groups are beginning to invest significant efforts into expanding the existing experimental strategies to study biomolecules in their native environment. One approach was recently summarized in a visionary paper by Sinz (179), who considered the possibilities provided by chemical cross-linking with MS detection for investigation of protein interactions in living cells. The idea of this approach is relatively straightforward and relies on moving the chemical reaction inside the cell. Once the proteins are cross-linked *in situ*, they can be isolated and analyzed by MS and MS/MS. In addition to simply identifying the interaction partners and mapping the interaction networks, this method should be capable of generating more detailed information on these interactions at the molecular level, (e.g., by localizing the cross-linked sites and providing distance constraints), as discussed in Section 4.2. Eventually, such experiments would be expected to produce low-resolution structures of protein complexes in their native environment, a task that would likely be assisted by available *in vitro* data and

computational methods of analysis. Despite its conceptual simplicity, this approach faces a formidable challenge, namely, the need to efficiently deliver the cross-linking reagents to the cell without disrupting its normal functioning or indeed killing it altogether. This obstacle places a significant limitation on the number of reagents that can be used in such *in vivo* measurements. In fact, formaldehyde is presently the only reagent that appears to be suitable for such type of work because of its unique ability to permeate cellular membranes (180). Another viable approach to this problem involves reprogramming the genetic code of the cell, making it produce and incorporate amino acids with reactive side chains into proteins, which can be used as *in situ* chemical probes (181).

Finally, we thought it would not be inappropriate to mention once again exciting work published recently by Zhu et al. (67), who studied an integral membrane porin OmpF in live *E. coli* cells using hydroxyl radical footprinting (see Section 9.1.4 for technical details of these experiments). This work went beyond a simple comparison of the conclusions derived from the labeling experiments and the available crystal structure: The method was used to probe voltage gating of this porin under different conditions (*E. coli* was stimulated with either a low ionic strength solution or a low pH buffer). Remarkably, even though virtually no changes were seen in the exterior portion of the protein, the oxidation pattern deep inside the porin channel was found to be sensitive to the cell's environment, revealing the molecular mechanism of gating of the porin in response to external cues (67).

9.5.3. "Live" Macromolecules: Equilibrium Systems or Dissipative Structures?

Another implicit assumption that has not been candidly articulated in this chapter is that the *in vivo* behavior of biomolecules can be properly modeled and studied under equilibrium conditions. In fact, everywhere in this text we have considered studies of various properties of biological macromolecules under equilibrium conditions. The only exception was Chapter 6, where kinetic processes were considered, but even there such processes were almost always leading from one equilibrium situation to another (e.g., protein refolding induced by changing the conditions from denaturing to near-native).

On the occasion of the centennial anniversary of the Gibbs papers on chemical equilibrium (182,183), the appropriateness of equilibrium conditions in the studies of living matter was first explicitly articulated by Tanford, who wrote: "*The time now seems appropriate to use his method of thermodynamic analysis as a basis of thinking about the organization of the living cell and, beyond that, the assembly of cells into multicellular organisms*" (184). Tanford suggested that understanding of the organization of living matter can be achieved by dividing the process into two parts: biosynthesis and assembly of biosynthetic products into organized structure. The latter process should be under thermodynamic control, (i.e., self-organized structures represent the macromolecular arrangement providing the lowest chemical potential, which, in the absence of chemical transformations, is equal to the free energy). According to this view, the cell can be thought of as a sophisticated reaction vessel, whose dynamics are governed by diffusion and random collisions of various chemical species. However, recent decades have brought the realization that most cellular processes are far from equilibrium: Live cells are polar structures, and their interior is neither homogeneous nor isotropic, with most of the cellular functions involving directional movement and transport (185). Furthermore, every living cell is essentially a nonequilibrium system, as it relies on the constant flow of matter and energy.

The philosophical and theoretical framework for understanding how such complicated systems can assemble and function under nonequilibrium conditions has been provided by work of the Brussels School of Physics (186–188). It resulted in the creation of nonlinear and nonequilibrium thermodynamics, which rationalizes the emergence and evolution of *far-from-equilibrium states* (189,190). These states often achieve very high levels of order, leading to *self-organization*, a phenomenon that is central to the assembly and functioning of the living cell and its individual compartments (191). Self-organization is often contrasted to self-assembly; the latter is viewed as association of macromolecules into an equilibrium structure, while the former is a steady-state (nonequilibrium) process. Self-organization is obviously essential for understanding how sophisticated biological machinery works inside and outside the cells, and its appreciation and thorough understanding seems to be a critical step, which is yet to be made, toward realization of Drexler's bold vision of creating functional nanorobots (192). The importance of nonequilibrium dynamics in biology can be easily appreciated by recalling two examples of macromolecular assemblies considered earlier in Sections 9.4.1 and 9.4.2, namely, ribosomes and proteasomes. These two examples frequently act in the cell simultaneously, but produce totally opposite outcomes. Obviously, there are myriads of other examples of high-fidelity monodirectional biological processes, which overcome the randomizing effect of Brownian motion by converting chemical energy into mechanical work. As was the case 7 years ago when we wrote the first edition of this book, there are still very few examples of using MS to advance our understanding of the nonequilibrium processes *in vivo*. Nevertheless, the tremendous potential of MS as a biophysical tool makes it certain that this technique will become a major player in removing the shroud of mystery from various nonequilibrium phenomena, thus bringing us closer to the understanding of life itself.

REFERENCES

1. Werten, P. J., Remigy, H. W., de Groot, B. L., Fotiadis, D., Philippsen, A., Stahlberg, H., Grubmuller, H., and Engel, A. (2002) Progress in the analysis of membrane protein structure and function, *FEBS Lett. 529*, 65–72.
2. Arora, A. and Tamm, L. K. (2001) Biophysical approaches to membrane protein structure determination, *Curr. Opin. Struct. Biol. 11*, 540–547.
3. Torres, J., Stevens, T. J., and Samso, M. (2003) Membrane proteins: the 'Wild West' of structural biology, *Trends Biochem. Sci. 28*, 137–144.
4. Garavito, R. M. and Ferguson-Miller, S. (2001) Detergents as tools in membrane biochemistry, *J. Biol. Chem. 276*, 32403–32406.
5. White, S. H., Ladokhin, A. S., Jayasinghe, S., and Hristova, K. (2001) How membranes shape protein structure, *J. Biol. Chem. 276*, 32395–32398.
6. Wu, C. C. and Yates, J. R., 3rd. (2003) The application of mass spectrometry to membrane proteomics, *Nat. Biotechnol. 21*, 262–267.
7. Sanders, C. R. and Oxenoid, K. (2000) Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins, *Biochim. Biophys. Acta 1508*, 129–145.
8. Bouchard, M., Benjamin, D. R., Tito, P., Robinson, C. V., and Dobson, C. M. (2000) Solvent effects on the conformation of the transmembrane peptide gramicidin A: insights from electrospray ionization mass spectrometry, *Biophys. J. 78*, 1010–1017.
9. Cross, T. A., Arseniev, A., Cornell, B. A., Davis, J. H., Killian, J. A., Koeppe, R. E., 2nd, Nicholson, L. K., Separovic, F., and Wallace, B. A. (1999) Gramicidin channel controversy - revisited, *Nat. Struct. Biol. 6*, 610–611.
10. Burkhart, B. M. and Duax, W. L. (1999) Gramicidin channel controversy—reply, *Nat. Struct. Biol. 6*, 611–612.
11. Chitta, R. K. and Gross, M. L. (2004) Electrospray ionization-mass spectrometry and tandem mass spectrometry reveal self-association and metal-ion binding of hydrophobic peptides: a study of the gramicidin dimer, *Biophys. J. 86*, 473–479.
12. Waring, A. J., Mobley, P. W., and Gordon, L. M. (1998) Conformational mapping of a viral fusion peptide in structure-promoting solvents using circular dichroism and electrospray mass spectrometry, *Proteins S2*, 38–49.
13. Poschner, B. C. and Langosch, D. (2009) Stabilization of conformationally dynamic helices by covalently attached acyl chains, *Protein Sci. 18*, 1801–1805.
14. Marsh, D. (2003) Lipid interactions with transmembrane proteins, *Cell. Mol. Life Sci. 60*, 1575–1580.
15. Warschawski, D. E., Arnold, A. A., Beaugrand, M., Gravel, A., Chartrand, É., and Marcotte, I. (2011) Choosing membrane mimetics for NMR structural studies of transmembrane proteins, *Biochim. Biophys. Acta 1808*, 1957–1974.
16. le Maire, M., Champeil, P., and Moller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents, *Biochim. Biophys. Acta 1508*, 86–111.
17. Linke, D. (2009) Detergents: an overview, In *Guide to Protein Purification, Second Edition* (Burgess, R. R., and Deutscher, M. P., Eds.), pp 603–617, Elsevier Academic Press Inc., San Diego.
18. Annesley, T. M. (2003) Ion suppression in mass spectrometry, *Clin. Chem. 49*, 1041–1044.
19. Fearnley, I. M. and Walker, J. E. (1996) Analysis of hydrophobic proteins and peptides by electrospray ionization MS, *Biochem. Soc. Trans. 24*, 912–917.
20. Hufnagel, P., Schweiger, U., Eckerskorn, C., and Oesterhelt, D. (1996) Electrospray ionization mass spectrometry of genetically and chemically modified bacteriorhodopsins, *Anal. Biochem. 243*, 46–54.
21. Whitelegge, J. P., Gundersen, C. B., and Faull, K. F. (1998) Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins, *Protein Sci. 7*, 1423–1430.
22. Whitelegge, J. P. (2004) HPLC and mass spectrometry of intrinsic membrane proteins, *Methods Mol. Biol. 251*, 323–340.
23. Barnidge, D. R., Dratz, E. A., Jesaitis, A. J., and Sunner, J. (1999) Extraction method for analysis of detergent-solubilized bacteriorhodopsin and hydrophobic peptides by electrospray ionization mass spectrometry, *Anal. Biochem. 269*, 1–9.
24. Ishihama, Y., Katayama, H., and Asakawa, N. (2000) Surfactants usable for electrospray ionization mass spectrometry, *Anal. Biochem. 287*, 45–54.
25. Cadene, M. and Chait, B. T. (2000) A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins, *Anal. Chem. 72*, 5655–5658.
26. Zhang, N., Doucette, A., and Li, L. (2001) Two-layer sample preparation method for MALDI mass spectrometric analysis of protein and peptide samples containing sodium dodecyl sulfate, *Anal. Chem. 73*, 2968–2975.
27. Zhang, N. and Li, L. (2002) Ammonium dodecyl sulfate as an alternative to sodium dodecyl sulfate for protein sample preparation with improved performance in MALDI mass spectrometry, *Anal. Chem. 74*, 1729–1736.
28. Lengqvist, J., Svensson, R., Evergren, E., Morgenstern, R., and Griffiths, W. J. (2004) Observation of an intact non-covalent homotrimer of detergent-solubilised rat microsomal glutathione transferase 1 by electrospray mass spectrometry, *J. Biol. Chem. 279*, 1311–13316.
29. Barrera, N. P., Di Bartolo, N., Booth, P. J., and Robinson, C. V. (2008) Micelles protect membrane complexes from solution to vacuum, *Science 321*, 243–246.
30. Kaltashov, I. A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal. Chem. 77*, 5370–5379.
31. Wang, S. C., Politis, A., Di Bartolo, N., Bavro, V. N., Tucker, S. J., Booth, P. J., Barrera, N. P., and Robinson, C. V. (2010) Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing, *J. Am. Chem. Soc. 132*, 15468–15470.
32. Charvat, A. and Abel, B. (2007) How to make big molecules fly out of liquid water: applications, features and physics of

laser assisted liquid phase dispersion mass spectrometry, *Phys. Chem. Chem. Phys. 9*, 3335–3360.

33. Meier, T., Morgner, N., Matthies, D., Pogoryelov, D., Keis, S., Cook, G. M., Dimroth, P., and Brutschy, B. (2007) A tridecameric c ring of the adenosine triphosphate (ATP) synthase from the thermoalkaliphilic Bacillus sp. strain TA2.A1 facilitates ATP synthesis at low electrochemical proton potential, *Mol. Microbiol. 65*, 1181–1192.
34. Hoffmann, J., Aslimovska, L., Bamann, C., Glaubitz, C., Bamberg, E., and Brutschy, B. (2010) Studying the stoichiometries of membrane proteins by mass spectrometry: microbial rhodopsins and a potassium ion channel, *Phys. Chem. Chem. Phys. 12*, 3480–3485.
35. Rey, M., Man, P., Clémençon, B., Trézéguet, V., Brandolin, G., Forest, E., and Pelosi, L. (2010) Conformational dynamics of the bovine mitochondrial ADP/ATP carrier isoform 1 revealed by hydrogen/deuterium exchange coupled to mass spectrometry, *J. Biol. Chem. 285*, 34981–34990.
36. Whitelegge, J. P., le Coutre, J., Lee, J. C., Engel, C. K., Prive, G. G., Faull, K. F., and Kaback, H. R. (1999) Toward the bilayer proteome, electrospray ionization-mass spectrometry of large, intact transmembrane proteins, *Proc. Natl. Acad. Sci. USA 96*, 10695–10698.
37. Weinglass, A. B., Whitelegge, J. P., Hu, Y., Verner, G. E., Faull, K. F., and Kaback, H. R. (2003) Elucidation of substrate binding interactions in a membrane transport protein by mass spectrometry, *EMBO J. 22*, 1467–1477.
38. Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R., and Iwata, S. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*, *Science 301*, 610–615.
39. Gupta, S., Bavro, V. N., D'Mello, R., Tucker, S. J., Vénien-Bryan, C., and Chance, M. R. (2010) Conformational changes during the gating of a potassium channel revealed by structural mass spectrometry, *Structure 18*, 839–846.
40. Buhler, S., Michels, J., Wendt, S., Ruck, A., Brdiczka, D., Welte, W., and Przybylski, M. (1998) Mass spectrometric mapping of ion channel proteins (porins) and identification of their supramolecular membrane assembly, *Proteins 2*, 63–73.
41. Pan, Y., Brown, L., and Konermann, L. (2009) Mapping the structure of an integral membrane protein under semi-denaturing conditions by laser-induced oxidative labeling and mass spectrometry, *J. Mol. Biol. 394*, 968–981.
42. Cross, T. A., Sharma, M., Yi, M., and Zhou, H.-X. (2011) Influence of solubilizing environments on membrane protein structures, *Trends Biochem. Sci. 36*, 117–125.
43. Blonder, J., Goshe, M. B., Moore, R. J., Pasa-Tolic, L., Masselon, C. D., Lipton, M. S., and Smith, R. D. (2002) Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry, *J. Proteome Res. 1*, 351–360.
44. Ball, L. E., Oatis, J. E., Dharmasiri, K., Busman, M., Wang, J., Cowden, L. B., Galijatovic, A., Chen, N., Crouch, R. K., and Knapp, D. R. (1998) Mass spectrometric analysis of integral membrane proteins: Application to complete mapping of bacteriorhodopsins and rhodopsin, *Protein Sci. 7*, 758–764.
45. Gelasco, A., Crouch, R. K., and Knapp, D. R. (2000) Intrahelical arrangement in the integral membrane protein rhodopsin investigated by site-specific chemical cleavage and mass spectrometry, *Biochemistry 39*, 4907–4914.
46. Dorman, G. and Prestwich, G. D. (1994) Benzophenone photophores in biochemistry, *Biochemistry 33*, 5661–5673.
47. Leite, J. F., Blanton, M. P., Shahgholi, M., Dougherty, D. A., and Lester, H. A. (2003) Conformation-dependent hydrophobic photolabeling of the nicotinic receptor: Electrophysiology-coordinated photochemistry and mass spectrometry, *Proc. Natl. Acad. Sci. USA 100*, 13054–13059.
48. Demmers, J. A., Haverkamp, J., Heck, A. J., Koeppe, R. E., 2nd, and Killian, J. A. (2000) Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers, *Proc. Natl. Acad. Sci. USA 97*, 3189–3194.
49. Demmers, J. A., van Duijn, E., Haverkamp, J., Greathouse, D. V., Koeppe, R. E., 2nd, Heck, A. J., and Killian, J. A. (2001) Interfacial positioning and stability of transmembrane peptides in lipid bilayers studied by combining hydrogen/deuterium exchange and mass spectrometry, *J. Biol. Chem. 276*, 34501–34508.
50. Demmers, J. A., Rijkers, D. T., Haverkamp, J., Killian, J. A., and Heck, A. J. (2002) Factors affecting gas-phase deuterium scrambling in peptide ions and their implications for protein structure determination, *J. Am. Chem. Soc. 124*, 11191–11198.
51. Marsh, D. (2002) Membrane water-penetration profiles from spin labels, *Eur. Biophys. J. 31*, 559–562.
52. Hansen, R. K., Broadhurst, R. W., Skelton, P. C., and Arkin, I. T. (2002) Hydrogen/deuterium exchange of hydrophobic peptides in model membranes by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom. 13*, 1376–1387.
53. Hebling, C. M., Morgan, C. R., Stafford, D. W., Jorgenson, J. W., Rand, K. D., and Engen, J. R. (2010) Conformational analysis of membrane proteins in phospholipid bilayer nanodiscs by hydrogen exchange mass spectrometry, *Anal. Chem. 82*, 5415–5419.
54. Cao, J., Hsu, Y.-H., Li, S., Woods, V. L., and Dennis, E. A. (2011) Lipoprotein-associated phospholipase A2 interacts with phospholipid vesicles via a surface-disposed hydrophobic α-helix, *Biochemistry 50*, 5314–5321.
55. Bogdanov, M. and Dowhan, W. (1999) Lipid-assisted protein folding, *J. Biol. Chem. 274*, 36827–36830.
56. Vereb, G., Szollosi, J., Matko, J., Nagy, P., Farkas, T., Vigh, L., Matyus, L., Waldmann, T. A., and Damjanovich, S. (2003) Dynamic, yet structured: The cell membrane three decades after the Singer–Nicolson model, *Proc. Natl. Acad. Sci. USA 100*, 8053–8058.
57. Clague, M. J. (1998) Molecular aspects of the endocytic pathway, *Biochem. J. 336*, 271–282.
58. Young, I., Wang, I., and Roof, W. D. (2000) Phages will out: strategies of host cell lysis, *Trends Microbiol. 8*, 120–128.
59. Bernhardt, T. G., Wang, I. N., Struck, D. K., and Young, R. (2002) Breaking free: "protein antibiotics" and phage lysis, *Res. Microbiol. 153*, 493–501.

60. Grundling, A., Manson, M. D., and Young, R. (2001) Holins kill without warning, *Proc. Natl. Acad. Sci. USA 98*, 9348–9352.
61. Cullen, B. R. (2001) Journey to the center of the cell, *Cell 105*, 697–700.
62. Poranen, M. M., Daugelavicius, R., and Bamford, D. H. (2002) Common principles in viral entry, *Annu. Rev. Microbiol. 56*, 521–538.
63. Akashi, S. and Takio, K. (2001) Structure of melittin bound to phospholipid micelles studied using hydrogen–deuterium exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *J. Am. Soc. Mass Spectrom. 12*, 1247–1253.
64. Hristova, K., Dempsey, C. E., and White, S. H. (2001) Structure, location, and lipid perturbations of melittin at the membrane interface, *Biophys. J. 80*, 801–811.
65. Pan, Y., Stocks, B. B., Brown, L., and Konermann, L. (2009) Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry, *Anal. Chem. 81*, 28–35.
66. Wen, J. Z., Zhang, H., Gross, M. L., and Blankenship, R. E. (2009) Membrane orientation of the FMO antenna protein from *Chlorobaculum tepidum* as determined by mass spectrometry-based footprinting, *Proc. Natl. Acad. Sci. USA 106*, 6134–6139.
67. Zhu, Y., Guo, T. N., Park, J. E., Li, X., Meng, W., Datta, A., Bern, M., Lim, S. K., and Sze, S. K. (2009) Elucidating *in vivo* structural dynamics in integral membrane protein by hydroxyl radical footprinting, *Mol. Cell. Proteomics 8*, 1999–2010.
68. Michaelis, L. (1902) Investigations on protein precipitation, *Dtsch. Med. Wochenschr. 28*, 733–736.
69. Jahn, T. R. and Radford, S. E. (2008) Folding versus aggregation: Polypeptide conformations on competing pathways, *Arch. Biochem. Biophys. 469*, 100–117.
70. Nettleton, E. J., Tito, P., Sunde, M., Bouchard, M., Dobson, C. M., and Robinson, C. V. (2000) Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry, *Biophys. J. 79*, 1053–1065.
71. Larson, J. L., Ko, E., and Miranker, A. D. (2000) Direct measurement of islet amyloid polypeptide fibrillogenesis by mass spectrometry, *Protein Sci. 9*, 427–431.
72. Villanueva, J., Villegas, V., Querol, E., Aviles, F. X., and Serrano, L. (2003) Monitoring disappearance of monomers and generation of resistance to proteolysis during the formation of the activation domain of human procarboxypeptidase A2 (ADA2h) amyloid fibrils by matrix assisted laser desorption ionization time-of-flight MS, *Biochem. J. 374*, 489–495.
73. Monti, M., Garolla di Bard, B. L., Calloni, G., Chiti, F., Amoresano, A., Ramponi, G., and Pucci, P. (2004) The regions of the sequence most exposed to the solvent within the amyloidogenic state of a protein initiate the aggregation process, *J. Mol. Biol. 336*, 253–262.
74. Kheterpal, I., Williams, A., Murphy, C., Bledsoe, B., and Wetzel, R. (2001) Structural features of the Aβ amyloid fibril elucidated by limited proteolysis, *Biochem 40*, 11757–11767.
75. Monti, M., Principe, S., Giorgetti, S., Mangione, P., Merlini, G., Clark, A., Bellotti, V., Amoresano, A., and Pucci, P. (2002) Topological investigation of amyloid fibrils obtained from β2-microglobulin, *Protein Sci. 11*, 2362–2369.
76. Haas, J., Vöhringer-Martinez, E., Bögehold, A., Matthes, D., Hensen, U., Pelah, A., Abel, B., and Grubmüller, H. (2009) Primary steps of pH-dependent insulin aggregation kinetics are governed by conformational flexibility, *Chembiochem 10*, 1816–1822.
77. Bartolini, M., Naldi, M., Fiori, J., Valle, F., Biscarini, F., Nicolau, D. V., and Andrisano, V. (2011) Kinetic characterization of amyloid-beta 1–42 aggregation with a multimethodological approach, *Anal. Biochem. 414*, 215–225.
78. Smith, A. M., Jahn, T. R., Ashcroft, A. E., and Radford, S. E. (2006) Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry, *J. Mol. Biol. 364*, 9–19.
79. Gruendling, T., Guilhaus, M., and Barner-Kowollik, C. (2009) Fast and accurate determination of absolute individual molecular weight distributions from mixtures of polymers via size exclusion chromatography-electrospray ionization mass spectrometry, *Macromolecules 42*, 6366–6374.
80. Kükrer, B., Filipe, V., van Duijn, E., Kasper, P., Vreeken, R., Heck, A., and Jiskoot, W. (2010) Mass spectrometric analysis of intact human monoclonal antibody aggregates fractionated by size-exclusion chromatography, *Pharm. Res. 27*, 2197–2204.
81. Wang, G., Johnson, A. J. and Kaltashov, I. A. (2011) Evaluation of electrospray ionization mass spectrometry as a tool for characterization of small soluble protein aggregates. *Anal. Chem.* (in press, DOI: 10.1021/ac203017x).
82. Abzalimov, R. R. and Kaltashov, I. A. (2010) Electrospray ionization mass spectrometry of highly heterogeneous protein systems: Protein ion charge state assignment via incomplete charge reduction, *Anal. Chem. 82*, 7523–7526.
83. Wang, G., Abzalimov, R. R., and Kaltashov, I. A. (2011) Direct monitoring of heat-stressed biopolymers with temperature-controlled electrospray ionization mass spectrometry, *Anal. Chem. 83*, 2870–2876.
84. Wang, G. and Kaltashov, I. A. (2011) Aggregation of antithrombin-III under heat-stress conditions proceeds via small-scale conformational change, (Manuscript in preparation.)
85. Kheterpal, I., Cook, K. D., Wetzel, R., Indu, K., and Ronald, W. (2006) Hydrogen/deuterium exchange mass spectrometry analysis of protein aggregates, *Methods Enzymol.*, *413*, 140–166.
86. Kheterpal, I. and Wetzel, R. (2006) Hydrogen/deuterium exchange mass spectrometry–a window into amyloid structure, *Acc. Chem. Res. 39*, 584–593.
87. Tito, P., Nettleton, E. J., and Robinson, C. V. (2000) Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: a site-specific investigation by mass spectrometry, *J. Mol. Biol. 303*, 267–278.
88. Nazabal, A., Dos Reis, S., Bonneu, M., Saupe, S. J., and Schmitter, J. M. (2003) Conformational transition occurring upon amyloid aggregation of the HET-s prion protein of *Podospora anserina* analyzed by hydrogen/deuterium exchange and mass spectrometry, *Biochemistry 42*, 8852–8861.

89. Kheterpal, I., Lashuel, H. A., Hartley, D. M., Walz, T., Lansbury, P. T., Jr., and Wetzel, R. (2003) Aβ protofibrils possess a stable core structure resistant to hydrogen exchange, *Biochemistry 42*, 14092–14098.
90. Wang, S. S., Tobler, S. A., Good, T. A., and Fernandez, E. J. (2003) Hydrogen exchange-mass spectrometry analysis of β-amyloid peptide structure, *Biochemistry 42*, 9507–9514.
91. Zhang, A., Qi, W., Good, T. A., and Fernandez, E. J. (2009) Structural differences between Aβ(1–40) intermediate oligomers and fibrils elucidated by proteolytic fragmentation and hydrogen/deuterium exchange, *Biophys. J. 96*, 1091–1104.
92. Ramos, I., Fabris, D., Qi, W., Fernandez, E. J., and Good, T. A. (2009) Kinetic study of β-amyloid residue accessibility using reductive alkylation and mass spectrometry, *Biotechnol. Bioeng. 104*, 181–192.
93. Sanchez, L., Madurga, S., Pukala, T., Vilaseca, M., Lopez-Iglesias, C., Robinson, C. V., Giralt, E., and Carulla, N. (2011) Aβ40 and Aβ42 amyloid fibrils exhibit distinct molecular recycling properties, *J. Am. Chem. Soc. 133*, 6505–6508.
94. Lu, X., Wintrode, P. L., and Surewicz, W. K. (2007) β-Sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange, *Proc. Nat. Acad. Sci. 104*, 1510–1515.
95. Damo, S. M., Phillips, A. H., Young, A. L., Li, S., Woods, V. L., and Wemmer, D. E. (2010) Probing the conformation of a prion protein fibril with hydrogen exchange, *J. Biol. Chem. 285*, 32303–32311.
96. Del Mar, C., Greenbaum, E. A., Mayne, L., Englander, S. W., and Woods, V. L., Jr., (2005) Structure and properties of α-synuclein and other amyloids determined at the amino acid level, *Proc. Natl. Acad. Sci. USA 102*, 15477–15482.
97. Kraus, M., Bienert, M., and Krause, E. (2003) Hydrogen exchange studies on Alzheimer's amyloid-beta peptides by mass spectrometry using matrix-assisted laser desorption/ionization and electrospray ionization, *Rapid Commun. Mass Spectrom. 17*, 222–228.
98. Pan, J., Han, J., Borchers, C. H., and Konermann, L. (2011) Conformer-specific hydrogen exchange analysis of Aβ(1–42) oligomers by top-down electron capture dissociation mass spectrometry, *Anal. Chem. 83*, 5386–5393.
99. Carulla, N., Zhou, M., Arimon, M., Gaira, M., Giralt, E., Robinson, C. V., and Dobson, C. M. (2009) Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation, *Proc. Natl. Acad. Sci. USA 106*, 7828–7833.
100. Hodkinson, J., Jahn, T., Radford, S., and Ashcroft, A. (2009) HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein β2-microglobulin upon release from the MHC-1, *J. Am. Soc. Mass. Spectrom. 20*, 278–286.
101. Zhang, A., Qi, W., Singh, S., and Fernandez, E. (2011) A new approach to explore the impact of freeze-thaw cycling on protein structure: hydrogen/deuterium exchange mass spectrometry (HX-MS), *Pharm. Res. 28*, 1179–1193.
102. Carulla, N., Zhou, M., Giralt, E., Robinson, C. V., and Dobson, C. M. (2010) Structure and intermolecular dynamics of aggregates populated during amyloid fibril formation studied by hydrogen/deuterium exchange, *Acc. Chem. Res. 43*, 1072–1079.
103. Egnaczyk, G. F., Greis, K. D., Stimson, E. R., and Maggio, J. E. (2001) Photoaffinity cross-linking of Alzheimer's disease amyloid fibrils reveals interstrand contact regions between assembled β-amyloid peptide subunits, *Biochemistry 40*, 11706–11714.
104. Bitan, G., Indu, K., and Ronald, W. (2006) Structural study of metastable amyloidogenic protein oligomers by photo-induced cross-linking of unmodified proteins, *Methods Enzymol.*, *413*, 217–236,
105. Piening, N., Weber, P., Hogen, T., Beekes, M., Kretzschmar, H., and Giese, A. (2006) Photo-induced crosslinking of prion protein oligomers and prions, *Amyloid 13*, 67–77.
106. Maji, S. K., Loo, R. R. O., Inayathullah, M., Spring, S. M., Vollers, S. S., Condron, M. M., Bitan, G., Loo, J. A., and Teplow, D. B. (2009) Amino acid position-specific contributions to amyloid beta-protein oligomerization, *J. Biol. Chem. 284*, 23580–23591.
107. Onisko, B., Fernandez, E. G., Freire, M. L., Schwarz, A., Baier, M., Camina, F., Garcia, J. R., Rodriguez-Segade Villamarin, S., and Requena, J. R. (2005) Probing PrPSc structure using chemical cross-linking and mass spectrometry: evidence of the proximity of Gly90 amino termini in the PrP 27–30 aggregate, *Biochem 44*, 10100–10109.
108. Mendoza, V. L., Antwi, K., Barón-Rodríguez, M. A., Blanco, C., and Vachet, R. W. (2010) Structure of the preamyloid dimer of β-2-microglobulin from covalent labeling and mass spectrometry, *Biochemistry 49*, 1522–1532.
109. Mendoza, V. L., Baron-Rodrieguez, M. A., Blanco, C., and Vachet, R. W. (2011) Structural insights into the pre-amyloid tetramer of β2-microglobulin from covalent labeling and mass spectrometry, *Biochemistry 50*, 6711–6722.
110. Bleiholder, C., Dupuis, N. F., Wyttenbach, T., and Bowers, M. T. (2011) Ion mobility-mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation, *Nat. Chem. 3*, 172–177.
111. Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B., and Bowers, M. T. (2005) Amyloid β-protein: monomer structure and early aggregation states of Aβ42 and its Pro^{19} alloform, *J. Am. Chem. Soc. 127*, 2075–2084.
112. Murray, M. M., Bernstein, S. L., Nyugen, V., Condron, M. M., Teplow, D. B., and Bowers, M. T. (2009) Amyloid β protein: β40 Inhibits β42 oligomerization, *J. Am. Chem. Soc. 131*, 6316–6317.
113. Iuraşcu, M., Cozma, C., Tomczyk, N., Rontree, J., Desor, M., Drescher, M., and Przybylski, M. (2009) Structural characterization of β-amyloid oligomer-aggregates by ion mobility mass spectrometry and electron spin resonance spectroscopy, *Anal. Bioanal. Chem. 395*, 2509–2519.
114. Abzalimov, R. R. and Kaltashov, I. A. (2011) Studies of conformational dynamics in highly heterogeneous systems by analysis of totally unresolved charge state distributions in ESI MS, (Manuscript in preparation).

115. Dobo, A. and Kaltashov, I. A. (2001) Detection of multiple protein conformational ensembles in solution via deconvolution of charge state distributions in ESI MS., *Anal. Chem. 73*, 4763–4773.

116. Mohimen, A., Dobo, A., Hoerner, J. K., and Kaltashov, I. A. (2003) A chemometric approach to detection and characterization of multiple protein conformers in solution using electrospray ionization mass spectrometry., *Anal. Chem. 75*, 4139–4147.

117. Houde, D., Berkowitz, S. A., and Engen, J. R. (2011) The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies, *J. Pharm. Sci. 100*, 2071–2086.

118. Fenn, L. S. and McLean, J. A. (2011) Structural resolution of carbohydrate positional and structural isomers based on gas-phase ion mobility-mass spectrometry, *Phys. Chem. Chem. Phys. 13*, 2196–2205.

119. Trimpin, S., Plasencia, M., Isailovic, D., and Clemmer, D. E. (2007) Resolving oligomers from fully grown polymers with IMS–MS, *Anal. Chem. 79*, 7965–7974.

120. Damen, C., Chen, W., Chakraborty, A., van Oosterhout, M., Mazzeo, J., Gebler, J., Schellens, J., Rosing, H., and Beijnen, J. (2009) Electrospray ionization quadrupole ion-mobility time-of-flight mass spectrometry as a tool to distinguish the lot-to-lot heterogeneity in N-glycosylation profile of the therapeutic monoclonal antibody trastuzumab, *J. Am. Soc. Mass Spectrom. 20*, 2021–2033.

121. Bagal, D., Zhang, H., and Schnier, P. D. (2008) Gas-phase proton-transfer chemistry coupled with TOF mass spectrometry and ion mobility-MS for the facile analysis of poly (ethylene glycols) and PEGylated polypeptide conjugates, *Anal. Chem. 80*, 2408–2418.

122. Mello, C. C. and Conte, D., Jr. (2004) Revealing the world of RNA interference, *Nature (London) 431*, 338–342.

123. Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell 116*, 281–297.

124. Lander, A. D. (1998) Proteoglycans: master regulators of molecular encounter? *Matrix Biol. 17*, 465–472.

125. Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L., and Chernushevich, I. V. (2005) Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex, *J. Am. Soc. Mass Spectrom. 16*, 998–1008.

126. Sharon, M., Witt, S., Felderer, K., Rockel, B., Baumeister, W., and Robinson, C. V. (2006) 20S proteasomes have the potential to keep substrates in store for continual degradation, *J. Biol. Chem. 281*, 9569–9575.

127. Moore, P. B. (2001) The ribosome at atomic resolution, *Biochemistry 40*, 3243–3250.

128. Moore, P. B. and Steitz, T. A. (2003) The structural basis of large ribosomal subunit function, *Annu. Rev. Biochem. 72*, 813–850.

129. Kopylov, A. M. (2002) X-ray analysis of ribosomes: the static of the dynamic, *Biochemistry 67*, 372–382.

130. Mathews, M. B. (2002) Lost in translation, *Trends Biochem. Sci. 27*, 267–269.

131. Preiss, T. and M, W. H. (2003) Starting the protein synthesis machine: eukaryotic translation initiation, *Bioessays 25*, 1201–1211.

132. Gabashvili, I. S., Agrawal, R. K., Spahn, C. M., Grassucci, R. A., Svergun, D. I., Frank, J., and Penczek, P. (2000) Solution structure of the E. coli 70S ribosome at 11.5 A resolution, *Cell 100*, 537–549.

133. Frank, J. (2003) Electron microscopy of functional ribosome complexes, *Biopolymers 68*, 223–233.

134. Benjamin, D. R., Robinson, C. V., Hendrick, J. P., Hartl, F. U., and Dobson, C. M. (1998) Mass spectrometry of ribosomes and ribosomal subunits, *Proc. Natl. Acad. Sci. USA 95*, 7391–7395.

135. Markus, M. A., Hinck, A. P., Huang, S., Draper, D. E., and Torchia, D. A. (1997) High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA, *Nat. Struct. Biol. 4*, 70–77.

136. Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M., and Robinson, C. V. (2000) Detection and selective dissociation of intact ribosomes in a mass spectrometer., *Proc. Natl. Acad. Sci. USA 97*, 5185–5190.

137. Hanson, C. L., Fucini, P., Ilag, L. L., Nierhaus, K. H., and Robinson, C. V. (2003) Dissociation of intact Escherichia coli ribosomes in a mass spectrometer. Evidence for conformational change in a ribosome elongation factor G complex, *J. Biol. Chem. 278*, 1259–1267.

138. Damoc, E., Fraser, C. S., Zhou, M., Videler, H., Mayeur, G. L., Hershey, J. W. B., Doudna, J. A., Robinson, C. V., and Leary, J. A. (2007) Structural characterization of the human eukaryotic initiation factor 3 protein complex by mass spectrometry, *Mol. Cell. Proteom. 6*, 1135–1146.

139. Gordiyenko, Y., Videler, H., Zhou, M., McKay, A. R., Fucini, P., Biegel, E., Muller, V., and Robinson, C. V. (2010) Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree, *Mol. Cell. Proteom. 9*, 1774–1783.

140. Gordiyenko, Y. and Robinson, C. V. (2008) The emerging role of MS in structure elucidation of protein–nucleic acid complexes, *Biochem. Soc. Trans. 36*, 723–731.

141. Beardsley, R. L., Running, W. E., and Reilly, J. P. (2006) Probing the structure of the *Caulobacter crescentus* ribosome with chemical labeling and mass spectrometry, *J. Proteome Res. 5*, 2935–2946.

142. Gruschke, S., Grone, K., Heublein, M., Holz, S., Israel, L., Imhof, A., Herrmann, J. M., and Ott, M. (2010) Proteins at the polypeptide tunnel exit of the yeast mitochondrial ribosome, *J. Biol. Chem. 285*, 19022–19028.

143. Yamamoto, T., Izumi, S., and Gekko, K. (2006) Mass spectrometry of hydrogen/deuterium exchange in 70S ribosomal proteins from *E. coli*, *FEBS Lett. 580*, 3638–3642.

144. Yamamoto, T., Shimizu, Y., Ueda, T., and Shiro, Y. (2010) Mg^{2+} dependence of 70 S ribosomal protein flexibility revealed by hydrogen/deuterium exchange and mass spectrometry, *J. Biol. Chem. 285*, 5646–5652.

145. Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S. X., Braig, K., Saibil, H. R., Fenton, W. A., and Horwich, A. L. (1995) Mechanism of GroEL action - productive release of polypeptide from a sequestered position under GroES, *Cell 83*, 577–587.

146. Robinson, C. V., Groß, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M., and Radford, S. E. (1994) Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry, *Nature (London) 372*, 646–651.

147. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding, *FEBS Lett. 262*, 20–24.

148. Robinson, C. V., Groß, M., and Radford, S. E. (1998) Probing conformations of GroEL-bound substrate proteins by mass spectrometry, *Methods Enzymol. 290*, 296–313.

149. Groß, M., Robinson, C. V., Mayhew, M., Hartl, F. U., and Radford, S. E. (1996) Significant hydrogen exchange protection in GroEL-bound DHFR is maintained during iterative rounds of substrate cycling, *Protein Sci. 5*, 2506–2513.

150. Goldberg, M. S., Zhang, J., Sondek, S., Matthews, C. R., Fox, R. O., and Horwich, A. L. (1997) Native-like structure of a protein-folding intermediate bound to the chaperonin GroEL, *Proc. Natl. Acad. Sci. USA 94*, 1080–1085.

151. Coyle, J. E., Texter, F. L., Ashcroft, A. E., Masselos, D., Robinson, C. V., and Radford, S. E. (1999) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism, *Nat. Struct. Biol. 6*, 683–690.

152. Gervasoni, P., Staudenmann, W., James, P., Gehrig, P., and Plückthun, A. (1996) beta-Lactamase binds to GroEL in a conformation highly protected against hydrogen/deuterium exchange, *Proc. Natl. Acad. Sci. USA 93*, 12189–12194.

153. Gervasoni, P., Staudenmann, W., James, P., and Plückthun, A. (1998) Identification of the binding surface on beta-lactamase for GroEL by limited proteolysis and MALDI mass spectrometry, *Biochemistry 37*, 11660–11669.

154. Gervasoni, P., Gehrig, P., and Plückthun, A. (1998) Two conformational states of beta-lactamase bound to GroEL: A biophysical characterization, *J. Mol. Biol. 275*, 663–675.

155. Chen, J. W., Walter, S., Horwich, A. L., and Smith, D. L. (2001) Folding of malate dehydrogenase inside the GroEL–GroES cavity, *Nat. Struct. Biol. 8*, 721–728.

156. van Duijn, E., Bakkes, P. J., Heeren, R. M. A., van den Heuvel, R. H. H., van Heerikhuizen, H., van der Vies, S. M., and Heck, A. J. R. (2005) Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry, *Nat. Methods 2*, 371–376.

157. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of the cell, *Nature (London) 450*, 973–982.

158. Minton, A. P. (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media, *J. Biol. Chem. 276*, 10577–10580.

159. Minton, A. P. (2000) Implications of macromolecular crowding for protein assembly, *Curr. Opin. Struct. Biol. 10*, 34–39.

160. Davis-Searles, P. R., Saunders, A. J., Erie, D. A., Winzor, D. J., and Pielak, G. J. (2001) Interpreting the effects of small uncharged solutes on protein-folding equilibria, *Annu. Rev. Biophys. Biomol. Struct. 30*, 271–306.

161. Hall, D. and Minton, A. P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges, *Biochim. Biophys. Acta 1649*, 127–139.

162. Medalia, O., Weber, I., Frangakis, A. S., Nicastro, D., Gerisch, G., and Baumeister, W. (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography, *Science 298*, 1209–1213.

163. Ellis, R. J. (2001) Macromolecular crowding: obvious but underappreciated, *Trends Biochem. Sci. 26*, 597–604.

164. Juraschek, R., Dulcks, T., and Karas, M. (1999) Nanoelectrospray - more than just a minimized-flow electrospray ionization source, *J. Am. Soc. Mass Spectrom. 10*, 300–308.

165. Bruggeman, F. J., van Heeswijk, W. C., Boogerd, F. C., and Westerhoff, H. V. (2000) Macromolecular intelligence in microorganisms, *Biol. Chem. 381*, 965–972.

166. Huang, S. and Ingber, D. E. (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks, *Exp. Cell Res. 261*, 91–103.

167. Pogun, S. (2001) Are attractors 'strange', or is life more complicated than the simple laws of physics? *Biosystems 63*, 101–114.

168. Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999) From molecular to modular cell biology, *Nature (London) 402*, C47–52.

169. Eisenberg, D., Marcotte, E. M., Xenarios, I., and Yeates, T. O. (2000) Protein function in the post-genomic era, *Nature (London) 405*, 823–826.

170. Kitano, H. (2002) Systems biology: a brief overview, *Science 295*, 1662–1664.

171. Gavin, A. C., and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man, *Curr. Opin. Chem. Biol. 7*, 21–27.

172. Collins, M. O. and Choudhary, J. S. (2008) Mapping multiprotein complexes by affinity purification and mass spectrometry, *Curr. Opin. Biotechnol. 19*, 324–330.

173. Monti, M., Cozzolino, M., Cozzolino, F., Vitiello, G., Tedesco, R., Flagiello, A., and Pucci, P. (2009) Puzzle of protein complexes *in vivo*: a present and future challenge for functional proteomics, *Exp. Rev. Proteom. 6*, 159–169.

174. Terentiev, A. A., Moldogazieva, N. T., and Shaitan, K. V. (2009) Dynamic proteomics in modeling of the living cell. Protein–protein interactions, *Biochem (Mosc.) 74*, 1586–1607.

175. Malik, R., Dulla, K., Nigg, E. A., and Korner, R. (2010) From proteome lists to biological impact- tools and strategies for the analysis of large MS data sets, *Proteomics 10*, 1270–1283.

176. Zhou, M. and Robinson, C. V. (2010) When proteomics meets structural biology, *Trends Biochem. Sci. 35*, 522–529.

177. Gavin, A. C., Maeda, K., and Kuhner, S. (2011) Recent advances in charting protein-protein interaction: mass

spectrometry-based approaches, *Curr. Opin. Biotechnol. 22*, 42–49.

178. Sardiu, M. E. and Washburn, M. P. (2011) Building protein–protein interaction networks with proteomics and informatics tools, *J. Biol. Chem. 286*, 23645–23651.
179. Sinz, A. (2010) Investigation of protein–protein interactions in living cells by chemical crosslinking and mass spectrometry, *Anal. Bioanal. Chem. 397*, 3433–3440.
180. Sutherland, B. W., Toews, J., and Kast, J. (2008) Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions, *J. Mass Spectrom. 43*, 699–715.
181. Xie, J. and Schultz, P. G. (2006) A chemical toolkit for proteins - an expanded genetic code, *Nat. Rev. Mol. Cell. Biol. 7*, 775–782.
182. Gibbs, J. W. (1876) On the equilibrium of heterogeneous substances, *Trans. Conn. Acad. 3*, 108–248.
183. Gibbs, J. W. (1878) On the equilibrium of heterogeneous substances, *Trans. Conn. Acad. 3*, 343–524.
184. Tanford, C. (1978) Hydrophobic effect and organization of living matter, *Science 200*, 1012–1018.
185. Bustamante, C., Keller, D., and Oster, G. (2001) The physics of molecular motors, *Acc. Chem. Res. 34*, 412–420.
186. Prigogine, I. and Wiame, J. M. (1946) Biologie et thermodynamique des phenomenes irreversibles, *Experientia 2*, 451–453.
187. Prigogine, I. (1972) La thermodynamique de la vie, *Recherche 3*, 547–570.
188. Prigogine, I., Nicolis, G., and Babloyan, A. (1974) Nonequilibrium problems in biological phenomena, *Ann. N. Y. Acad. Sci. USA 231*, 99–105.
189. Nicolis, G. and Prigogine, I. (1977) *Self-organization in nonequilibrium systems: from dissipative structures to order through fluctuations*, John Wiley & Sons, New York.
190. Kondepudi, D. K. and Prigogine, I. (1998) *Modern thermodynamics: from heat engines to dissipative structures*, John Wiley & Sons, New York.
191. Misteli, T. (2001) The concept of self-organization in cellular architecture, *J. Cell Biol. 155*, 181–186.
192. Drexler, K. E. (2006) *Engines of Creation 2.0: The coming era of nanotechnology (20-th anniversary edition)*, WOWIO LLC, Los Angeles.

APPENDIX

PHYSICS OF ELECTROSPRAY

Electrospray ionization is a convoluted process that involves a large number of steps, each having a profound effect on the outcome of the measurements. It occurs upon forcing a continuous flow of conducting liquid through a metal* capillary (needle), which is maintained at a high electrostatic potential (as discussed in Chapter 4). Such conditions often cause liquid spray and form aerosol particles, a phenomenon that was explored in detail and photographed by Zeleny in the 1910s (1). The electrospray process was rediscovered in the early 1950s and was thoroughly investigated for several decades. Aerosol production may occur in a variety of functional modes, which have been reviewed by Cloupeau and Prunet-Foch (2). One of the most commonly observed modes of aerosol formation is a *dripping mode* (or "field enhanced dripping"), in which drops are formed at regular time intervals. The drops usually have a diameter exceeding that of a capillary, however, such "main drops" can be sometimes accompanied by formation of several satellite drops (2). In a *microdripping mode,* the emission also occurs with single droplets (as in the conventional dripping mode), but the droplets have a much smaller diameter compared to that of the capillary. A jet appears at the apex of the liquid meniscus, and liquid begins to accumulate at the end of the jet, finally resulting in the formation of a single droplet. Following the droplet separation and jet retraction, the cycle begins again (2). This regime can easily evolve into the *cone-jet mode*, which is usually favored at higher electric fields. The occurrence of this mode of electrospraying was first documented by Zeleny (1), who noted that low-viscosity liquids (e.g., alcohol) were capable of forming "a cone with a fine thread of liquid coming from its apex". This thread "*remains intact for but a short distance . . . breaking up into drops rather suddenly*". The drops spread "into a more or less conical volume", leading to the appearance of a "brush spray" (1). This mode appears to be quite robust and can be maintained within a wide range of conditions. Intermittent (pulsed) cone-jet is observed once the voltage at the capillary drops below a certain level. Further decrease of the electric field may result in transition to a multicone jet, which is characterized by the formation of two (or more) emissive cusps at the tip of the capillary. At very high fields, the permanent part of the spray may disappear, giving rise to a totally random spraying (2).

Analytical treatment of the cone-jet regime was first attempted by Taylor (3), who obtained an exact equilibrium solution for the shape of the cone in the absence of liquid emission (no jet formed). Under this "local equilibrium" (surface tension is balanced by electric forces) a cone is formed, whose opening angle is 98.6° (3). Taylor's exact electrostatic solution can be extended analytically to account for ejection of a jet from the cone apex† (4,5). Consideration of these analyses extends beyond the scope of this book, however, we will mention one conclusion that

*Although most ESI sources do have a metal or metal-coated capillary as a means of delivering the flow of liquid to the interface region, it is possible to use other types of capillaries as well (e.g., fused silica).

†The jet is modeled as being infinitely long, (i.e., its disintegration into a "brush spray" is ignored.)

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

is particularly important. It relates to field penetration into the liquid during the electrospray process. Even if the liquid is not perfectly conducting, its bulk will remain quasineutral given the motion is slow enough to allow for electric relaxation (4). In this case the electric field in the liquid bulk is very small compared to the external field.

Break-up of the charged jet into a "brush spray" can be better understood by first considering disintegration of electrically neutral jets of fluids. Rayleigh considered perturbation of the infinite cylindrical flow of liquid in the form of harmonic oscillation along the spray axis z (6):

$$r = r_o + \alpha e^{-ikz} \tag{A-1}$$

where r and r_o are the radii of perturbed and unperturbed jet surfaces, and $k = 2\pi/\lambda$ is a wavenumber of the disturbance. It was found that the most rapidly propagating perturbation corresponds to a wavelength λ, which is ~ 4.5 times the diameter of the jet [a more extensive discussion of this problem, including the influence of viscosity, can be found in a recent review by Eggers (7)]. This defines the geometrical scale of jet disintegration in the so-called *varicose mode*, with the expected droplet diameter being almost two times that of the jet. Surprisingly, this holds even for electrically charged jets (8).

The oscillations, of course, do not have to be axisymmetric, and the perturbations can be presented in the general form as (9–11):

$$r = r_o + \alpha e^{(\omega t + im\theta - ikz)} \tag{A-2}$$

where $m\theta$ represents an angular component and ω is the growth rate of perturbations. The varicose instability occurs when $m = 0$, otherwise the jet shape (and break-up) will be dependent on the angle θ. If $m = 1$, then there will be alternating regions of θ, where r is either larger or smaller than r_o. This represents a whipping* motion of the jet (9), a phenomenon often observed at higher electrical field strength and liquid flow rates. If $m = 2$, the jet is no longer circular. An extreme example of such a jet is a so-called *ramified jet*, which can arise due to an insufficient flow rate of liquid and high enough electric field. This leads to excessive charge on the jet, and the electric stresses overcome the surface tension (9). A succession of thickened regions is formed on the main jet, each giving rise to fine jets. All three perturbations considered here ($m = 0$, 1, and 2) can arise simultaneously, and the break-up of a ramified jet is a complex mixture of all three types of perturbations (9).

An external electric field will accelerate the jet, leading to an increase of the perturbation wavelength. If such acceleration is very fast, the jet break-up may never occur. In most cases, however, the break-up will occur somewhere along the jet, leading to the formation of a droplet whose diameter in the varicose regime can be estimated as (4,9):

$$d \sim Q^{1/2} \left(\frac{\rho \varepsilon_o}{\gamma K} \right)^{1/6} \tag{A-3}$$

where Q is the flow rate, ρ is the density of the liquid, γ is its surface tension coefficient, and K is its conductivity.† There is an apparent disagreement in the literature as to what the average charge on the droplet is. de la Mora suggested that the jet charge "remains nearly tied to the liquid during the break-up process", so that the charge density is nearly constant for a given spray (12):

$$q \sim d^3 \tag{A-4}$$

Others maintain that a droplet charge/diameter ratio should be

$$q \sim d^{3/2} \tag{A-5}$$

and present experimental evidence that the power function coefficient can actually be as low as 1.2 (9).

Formation of the main droplets is accompanied by the appearance of smaller secondary (satellite) droplets. The "satellite" droplets likely originate from the atomization of the liquid bridges connecting the incipient "main" droplet with the rest of the jet just prior to their separation. The frequency of droplet production is typically very high (it can easily break the MHz level), and their velocities may reach tens of meters per second (9). Electrostatic repulsion between the droplets results in their divergence from the jet axis to form a brush spray observed by Zeleny (1). If one ignores solvent evaporation from the droplets, as well as their deformation, the dynamics of droplet movement in the external field E_{ext} can be described simply as (10):

$$m_i \ddot{\vec{r}}_i = q_i \vec{E}_{ext}(\vec{r}_i) + \sum_{j \neq i} \frac{q_i q_j (\vec{r}_i - \vec{r}_j)}{4\pi\varepsilon_o |\vec{r}_i - \vec{r}_j|^3} + C_D \frac{\pi d_i^2}{8} \rho \left| \dot{\vec{r}}_i - \vec{u} \right| \cdot \left(\dot{\vec{r}}_i - \vec{u} \right) \tag{A-6}$$

where m_i, q_i, and r_i are the droplet's mass, net charge and position, respectively. The drag force coefficient, gas density, and flow rate are represented by C_D, ρ, and $\vec{u}$. The space-charge effect (represented in Eq. A-6 by the Coulombic term) is quite significant, and its influence would be particularly high in the case of smaller satellite droplets. Because of its lower inertia and drag force, a smaller droplet

*This type of jet instability is often termed lateral "kink-type" instability (2).

†Droplets formed in the presence of a "kink-type" instability are scaled as $d \sim Q^{1/3}$.

will attain larger radial velocity. As a result, droplets are segregated according to their size in the "brush" part of the spray, with the area along the jet axis dominated by large droplets, while the smaller ones are forced to the periphery of the brush (10).

Droplet fate and ion formation:charged residue model (CRM). The ultimate fate of the droplet is determined by two parallel processes: solvent loss due to evaporation and electrodynamic instability. Even an uncharged conducting droplet will not maintain an ideal spherical form in an external electrostatic field. Such a droplet is likely to assume the shape of an oblate spheroid and generate uneven charge distribution on its surface (maintaining zero net charge) to maintain field-free conditions in the bulk of the liquid (13). The dynamics of conducting liquid drops in electrostatic fields was first investigated experimentally by Macky (14). He observed that given sufficient field strength, the droplets deform (elongate) and become pointed at the ends. An increase of the surface tension will be compensated for by a decrease of electrostatic energy (elongation must be accomplished by charge segregation). A further increase of the field results in the rapid development of instability, with liquid jets being ejected from the cusps (14). It appears that this process can be adequately described using the Taylor formalism just discussed (3). The droplet does not have to carry a net charge to become unstable In fact, the original experiments were performed with electrically neutral droplets, however, the field strength required to induce such instability is very high, approaching 10 kV/cm (14). The presence of net positive (or negative) charge on the droplet is likely to reduce the instability threshold.

Perhaps a more efficient droplet atomization process is related to the so-called Rayleigh instability and does require external electric field (15). The electrostatic repulsion among the like charges on the droplet is offset by a cohesive action of the surface tension. The potential energy of the former is inversely proportional to the radius of the droplet, while the energy of the latter is proportional to the square of the radius (assuming the droplet has a spherical form). If solvent evaporation from the droplet proceeds without loss of charge, the electrostatic component of the potential energy will increase dramatically upon droplet shrinkage, while the stabilizing role of the surface tension will be diminished. Eventually, the droplet will become unstable with respect to high multipole oscillations, resulting in "*liquid . . . thrown out in fine jets, whose fineness, however, has a limit*" (15). The instability criterion is

$$\gamma = \frac{(Ne)^2}{8\pi\varepsilon_o d_o^3} \tag{A-7}$$

where Ne represents a net charge of the droplet and d_o its diameter in the spherical shape. Numerous measurements have confirmed that droplet fission does occur at or slightly below the Rayleigh limit (16). The characteristic time of instability development can be estimated as (17,18):

$$\tau \sim d_o^2 \sqrt{\frac{\rho}{\gamma}} \tag{A-8}$$

which is on the order of one-tenth of a second for a water droplet with a radius of ~ 1 mm. Since the "typical" droplet size in ESI is three orders of magnitude below that (19), the characteristic time for instability development would fall below a microsecond, significantly below the rate of droplet evaporation.

Although the original consideration of the droplet fission problem by Rayleigh predicts formations of fine jets at the droplet's ends, a notion confirmed experimentally by many groups, another outcome, termed *rough fission*, is also possible. In this case, the precursor droplet is fragmented into a few droplets of similar size. It seems, however, that such a mode of fission only occurs when the liquid is a poor conductor (20). Droplets composed of conducting liquids (e.g., water, alcohols, which make, of course, a more interesting case as typical ESI solvents) do decompose in a *fine fission mode* (21). Such fission removes a significant proportion of the net charge from the droplet (up to 25–33%), while the mass loss is negligible (on the order of 0.3% or even less) (16,22). The droplets generated during jet disintegration are of similar size and close to an instability criterion (Eq. A-7), which means they undergo fission soon after separation from the main droplet. In the meantime, the main droplet relaxes to a spherical form following the ejection of the jet, as its net charge falls comfortably below the instability limit (Eq. A-7).

de la Mora considered the "fine fission" of conducting droplets charged to a Rayleigh limit (Eq. A-7) in the absence of the electric field (21). He concluded that the fission occurs via formation of a Taylor cone, followed by ejection of a fine jet from its tip. Although the lifetime of a Taylor cone is small compared to the characteristic droplet evaporation time, the process can still be considered quasistatic in the case of conducting liquids (the electrical relaxation time for conducting liquids is much smaller than the expected lifetime of a Taylor cone). As a result, one can use Taylor's original formalism (3) to describe fission processes. Continuous evaporation brings the diameter of the initially stable droplet to a Rayleigh limit, resulting in a Coulombic explosion. The precursor droplet loses a significant proportion of its charge (but not volume), thus becoming stable again. Subsequent solvent evaporation can, of course, bring this droplet back to the unstable conditions, resulting in another fission event (Fig. A1).

Eventually, charges on both the precursor and the progeny droplets will be reduced to such levels that further

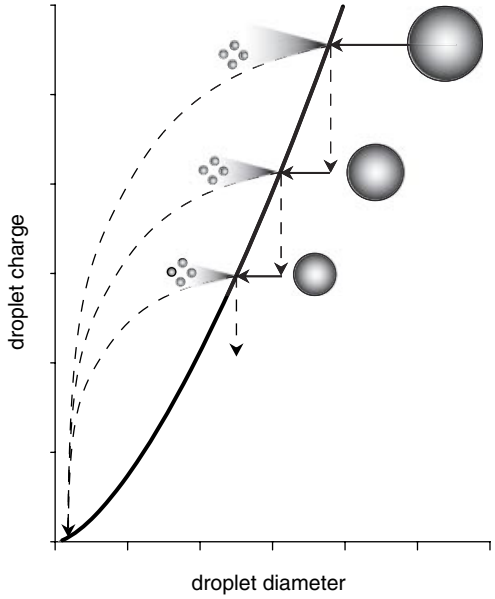


Figure A1. A schematic diagram representing evolution of a conducting charged droplet. Horizontal arrows indicate solvent loss due to evaporation prior to reaching the Rayleigh instability limit (solid curve). Dashed lines indicate ejection of progeny droplets (curved lines) and loss of charge–relaxation of the precursor droplet (straight vertical lines) following ejection of the progeny droplets.

solvent evaporation will no longer bring about Rayleigh instability. Complete evaporation of solvent from the droplets will, therefore, force the residual charge onto the solute molecules. The solute molecules are assumed to be large macromolecules that either lack charge in solution or else maintain electroneutrality through association with counterions. This scenario forms a foundation of the so-called CRM of ion formation, which is attributed to Dole et. al. (23). Once again, it should probably be mentioned that despite the violent nature of the charged jet break-up and subsequent Coulombic explosions of the electrospray droplets, the bulk of the liquid may remain essentially charge- and field-free (24). Given the sufficiently high concentration of charge carrier in the liquid, a quasiequilibrium charge layer will be formed at the liquid–gas interface. Consequently, even the weakly conducting liquid would conform to a quasilectrostatic model with the charge confined to the surface and the bulk of the liquid remaining quasineutral and field-free. This assertion, however, is not universally accepted (see, e.g., (12)). Therefore, the solute molecules residing in the bulk of the liquid will remain "oblivious" to the harsh conditions on the droplet surface and the environment beyond it until the last of the solvent is gone. Indeed, there is solid evidence that proteins do not denature during droplet evolution in the course of ESI (25). This conclusion is very important, since it opens up an opportunity to use ESI mass spectrometry (MS) to study biological macromolecules in their "near-native" environment.

Alternative view of ion formation:ion evaporation model (IEM). An alternative model of ion formation was offered by Iribarne and Thomson (26), who suggested that under certain conditions a single ion (or an ion cluster) can evaporate from the surface of the droplet. Removal of an ion from the droplet becomes thermodynamically favored as soon as the loss of solvation energy is offset by enthalpic gains due to the separation of the charges of the same polarity (electrostatic repulsion between the escaped ion and the remaining charges on the droplet). However, such an escape must proceed through an energy barrier created by the attraction between the ion and its electrical image on the surface of the droplet (26):

$$\Delta G^{\ddagger} = 4\pi\gamma r^2 + \frac{e^2}{2r}\left(1-\frac{1}{\varepsilon}\right) - e^2\left[\frac{N(\chi_{\mathrm{m}}+d)}{(R-d)(R+\chi_{\mathrm{m}})} + \frac{1}{4\chi_{\mathrm{m}}}\right] \tag{A-9}$$

where the first two terms represent the solvation energy of an ion of radius r in a liquid whose dielectric constant is ε and surface tension coefficient is γ. The last term represents the electrostatic interaction between the ion* and the droplet of radius R carrying N charges. The equilibrium position of the ion inside the droplet (prior to evaporation) is d, and the location of the barrier maximum (point of no return) is χ_{m}:

$$\chi_m = \frac{R}{2\sqrt{N}-1} \tag{A-10}$$

In order for the ion evaporation process to be efficient, it must occur faster than the total evaporation of solvent from the droplet, thus setting an upper limit for $\Delta G^{\ddagger}$ of ~ 9 kcal/mol in the case of water droplets (26). Substitution of this value in Eq. A-9 allows a critical droplet "radius–charge" relationship to be established. As one might expect, small ion evaporation is more efficient, compared to larger ones (Fig. A2). Importantly, droplets carrying relatively few charges have the critical size for ion evaporation exceeding that of the Rayleigh instability, indicating that there are conditions when the reduction of the droplet charge occurs through expulsion of a single ion, not through a fission process (Fig. A2).

Although the original Iribarne–Thomson work dealt with singly charged cluster ions, it was later invoked to explain production of large macromolecular ions as well (27,28). However, it is still unclear how much is contributed by the ion evaporation to the total macromolecular ion production. Reservations regarding the validity of IEM were first

*Evaporated ions are assumed to be singly charged; extension to multiply charged ions is rather straightforward.

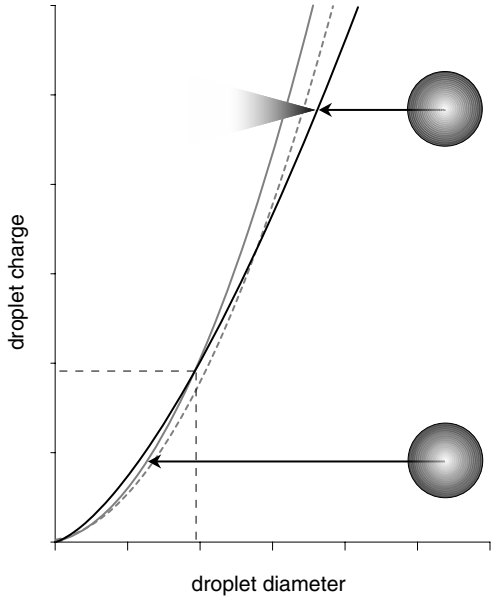


Figure A2. Two scenarios of a charged conducting droplet fate in a strong electrostatic field. Depending on the initial droplet charge, solvent loss due to evaporation will lead to either Coulombic explosion (Rayleigh instability limit is indicated with a solid black curve) or ion evaporation (a critical radius for a large ion evaporation as a function of the droplet total charge is indicated with a solid gray curve). Dashed gray curve indicates critical droplet radius for evaporation of a small ion from the droplet.

articulated by Röllgen, who questioned the use of classical electrostatics to describe processes that are essentially dynamic (29). Furthermore, it was suggested that "*the removal of an ion with part of its solvation sphere from a charged liquid surface under field stress represents the onset of an electrohydrodynamic disintegration process on a molecular scale*" (30). However, Labowsky, Fenn and de la Mora (31) recently pointed out that "*the process of emission of a singly charged nanodrop from a larger drop is radically different from a Coulomb explosion, being in fact indistinguishable from Iribarne–Thomson ionization*" (31). Such nanoscale processes (ion or nanodroplet desorption from the surface of the precursor droplet) should be facilitated by the discrete nature of the surface charge which is shown to increase the local electrostatic pressure by at least an order of magnitude compared to the continuous charge distribution (32). Still, even though the careful analytical treatment of such processes indicates that there is a distinct possibility of small ion cluster evaporation from the charged droplet, a window of conditions that lead to ion evaporations appears to be very narrow (31). In the case of water droplets, evaporation of small ions can only occur from relatively small droplets close to the Rayleigh instability limit. A possibility of larger ion (e.g., biopolymers) desorption from electrosprayed droplets still awaits careful consideration, although the accumulated experimental evidence suggests that production of such macro-ions occurs via Dole's CRM mechanism (19,33,34).

A very distinct feature of ESI is the accumulation of multiple charges on a single analyte molecule during ionization. Both CRM and IEM models predict accumulation of a significant number of charges on a macromolecule, and this number is expected to increase as the physical size of the molecule increases. According to the CRM model, the number of charges on the macromolecular ion will be determined by the total charge of the fission-incompetent droplet whose lower limit is set by the size of the macromolecule it encapsulates. Here we ignore gas phase processes that may result in charge transfer to/from the macromolecular ions. The IEM scenario of macromolecular ion desorption from a charged droplet invokes the notion of equidistant charges immobilized on the surface of the droplet, a configuration that minimizes the electrostatic repulsion energy (27). The number of charges accumulated by the macromolecular ion would then be determined by its cross-section upon traversing the surface of the droplet.

REFERENCES

1. Zeleny J. (1917) Instability of electrified liquid surfaces, *Phys. Rev. 10* 1–6.
2. Cloupeau M. and Prunet-Foch B. (1994) Electrohydrodynamic spraying functioning modes: A critical review, *J. Aerosol. Sci. 25* 1021–1036.
3. Taylor G. (1964) Disintegration of water drops in an electrical field, *Proc. Roy. Soc. London A 280* 383–397.
4. Ganan-Calvo A. M. (1997) Cone-jet analytical extension of Taylor's electrostatic solution and the asymptotic universal scaling laws in electrospraying, *Phys. Rev. Lett. 79*, 217–220.
5. Cherney L. T. (1999) Electrohydrodynamics of electrified liquid menisci and emitted jets, *J. Aerosol. Sci. 30*, 851–862.
6. Rayleigh J. W. S. (1878) On the instability of jets, *Proc. London Math. Soc. 11*, 4–13.
7. Eggers J. (1997) Nonlinear dynamics and breakup of free-surface flows. *Rev. Mod. Phys. 69*, 865–929.
8. Cloupeau M. and Prunet-Foch B. (1989) Electrostatic spraying of liquids in cone-jet mode, *J. Electrostat. 22*, 135–159.
9. Hartman R. P. A., Brunner D. J., Camelot D. M. A., Marijnissen J. C. M. and Scarlett B. (2000) Jet break-up in electrohydrodynamic atomization in the cone-jet mode, *J. Aerosol Sci. 31* 65–95.
10. Hartman R. P. A., Borra J. P., Brunner D. J., Marijnissen J. C. M., and Scarlett B. (1999) The evolution of electrohydrodynamic sprays produced in the cone-jet mode, a physical model, *J. Electrostat. 47*, 143–170.
11. Hartman R. P. A., Brunner D. J., Camelot D. M. A., Marijnissen J. C. M. and Scarlett B. (1999) Electrohydrodynamic atomization in the cone-jet mode physical modeling of the liquid cone and jet, *J. Aerosol Sci. 30*, 823–849.

12. de Juan L. and de la Mora J. F. (1997) Charge and size distributions of electrospray drops, *J. Colloid Interf. Sci. 186* 280–293.
13. Landau L. D., Lifshitz E. M., Pitaevskii L. P. (1984) *Electrodynamics of continuous media*. Elsevier Butterworth-Heinemann, Oxford.
14. Macky W. A. (1931) Some investigations on the deformation and breaking of water drops in strong electric fields, *Proc. R. Soc. London A 133* 565–587.
15. Rayleigh J. W. S. (1882). On the equilibrium of liquid conducting masses charged with electricity, *Philos. Mag. 14* 184–186.
16. Duft D., Lebius H., Huber B. A., Guet C. and Leisner T. (2002) Shape oscillations and stability of charged microdroplets, *Phys. Rev. Lett. 89*, art. no. 084503.
17. Belonozhko D. F. and Grigor'ev A. I. (1999) Characteristic time for the evolution of instability of a droplet charged to the rayleigh limit, *Tech. Phys. Lett. 25* 610–611.
18. Belonozhko D. F. and Grigor'ev A. I. (2000). Nonlinear capillary oscillation of a charged drop, *Tech. Phys. 45* 1001–1008.
19. Kebarle P. and Peschke M. (2000). On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions, *Analyt. Chim. Acta 406*, 11–35.
20. Richardson C. B., Pigg A. L. and Hightower R. L. (1989) On the stability limit of charged droplets, *Proc. R. Soc. London Ser. A 422*, 319–328.
21. de la Mora J. F. (1996) On the outcome of the Coulombic fission of a charged isolated drop, *J. Coll. Int. Sci. 178* 209–218.
22. Duft D., Achtzehn T., Muller R., Huber B. A. and Leisner T. (2003). Coulomb fission–Rayleigh jets from levitated microdroplets, *Nature* (London) *421*, 128–128.
23. Dole M., Mack L. L. and Hines R. L. (1968) Molecular beams of macroions, *J. Chem. Phys. 49* 2240–2249.
24. Ganan-Calvo A. M. (1999) The surface charge in electrospraying: Its nature and its universal scaling laws, *J. Aerosol Sci. 30* 863–872.
25. Mirza U. A. and Chait B. T. (1997) Do proteins denature during droplet evolution in electrospray ionization? *Int. J. Mass Spectrom. Ion Proc. 162* 173–181.
26. Iribarne J. V. and Thomson B. A. (1976) On the evaporation of small ions from charged droplets, *J. Chem. Phys. 64*, 2287–2294.
27. Fenn J. B. (1993) Ion formation from charged droplets: Roles of geometry, energy and time, *J. Am. Soc. Mass Spectrom. 4*, 524–535.
28. Fenn J. B., Rosell J. and Meng C. K. (1997) In electrospray ionization, how much pull does an ion need to escape its droplet prison? *J. Am. Soc. Mass Spectrom. 8*, 1147–1157.
29. Röllgen F. W., Bramer-Weger E. and Butfering L. (1987) Field-ion emission from liquid solutions - ion evaporation against electrohydrodynamic disintegration, *J. Phys. 48*, 253–256.
30. Rollgen F. W., Bramer-Weger E. and Butfering L. (1987) Field-ion emission from liquid solutions–ion evaporation against electrohydrodynamic disintegration, *J. Phys. 48*, 253–256.
31. Labowsky M., Fenn J. B. and de la Mora J. F. (2000) A continuum model for ion evaporation from a drop: Effect of curvature and charge on ion solvation energy, *Analyt. Chim. Acta 406*, 105–118.
32. Labowsky M. (1998) Discrete charge distributions in dielectric droplets, *J. Coll. Inter. Sci. 206*, 19–28.
33. de la Mora J. F. (2000) Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Analyt. Chim. Acta 406*, 93–104.
34. Kebarle P. (2000) A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry, *J. Mass Spectrom. 35*, 804–817.

INDEX

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

WILEY SERIES ON MASS SPECTROMETRY

Series Editors

Dominic M. Desiderio
Departments of Neurology and Biochemistry
University of Tennessee Health Science Center

Nico M. M. Nibbering
Vrije Universiteit Amsterdam, The Netherlands

John R. de Laeter • *Applications of Inorganic Mass Spectrometry*
Michael Kinter and Nicholas E. Sherman • *Protein Sequencing and Identification Using Tandem Mass Spectrometry*
Chhabil Dass • *Principles and Practice of Biological Mass Spectrometry*
Mike S. Lee • *LC/MS Applications in Drug Development*
Jerzy Silberring and Rolf Eckman • *Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research*
J. Wayne Rabalais • *Principles and Applications of Ion Scattering Spectrometry: Surface Chemical and Structural Analysis*
Mahmoud Hamdan and Pier Giorgio Righetti • *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis, and Microarray Technology*
Igor A. Kaltashov and Stephen J. Eyles • *Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition*
Isabella Dalle-Donne, Andrea Scaloni, and D. Allan Butterfield • *Redox Proteomics: From Protein Modifications to Cellular Dysfunction and Diseases*
Silas G. Villas-Boas, Ute Roessner, Michael A.E. Hansen, Jorn Smedsgaard, and Jens Nielsen • *Metabolome Analysis: An Introduction*
Mahmoud H. Hamdan • *Cancer Biomarkers: Analytical Techniques for Discovery*
Chabbil Dass • *Fundamentals of Contemporary Mass Spectrometry*
Kevin M. Downard (Editor) • *Mass Spectrometry of Protein Interactions*
Nobuhiro Takahashi and Toshiaki Isobe • *Proteomic Biology Using LC-MS: Large Scale Analysis of Cellular Dynamics and Function*
Agnieszka Kraj and Jerzy Silberring (Editors) • *Proteomics: Introduction to Methods and Applications*
Ganesh Kumar Agrawal and Randeep Rakwal (Editors) • *Plant Proteomics: Technologies, Strategies, and Applications*
Rolf Ekman, Jerzy Silberring, Ann M. Westman-Brinkmalm, and Agnieszka Kraj (Editors) • *Mass Spectrometry: Instrumentation, Interpretation, and Applications*
Christoph A. Schalley and Andreas Springer • *Mass Spectrometry and Gas-Phase Chemistry of Non-Covalent Complexes*
Riccardo Flamini and Pietro Traldi • *Mass Spectrometry in Grape and Wine Chemistry*
Mario Thevis • *Mass Spectrometry in Sports Drug Testing: Characterization of Prohibited Substances and Doping Control Analytical Assays*
Sara Castiglioni, Ettore Zuccato, and Roberto Fanelli • *Illicit Drugs in the Environment: Occurrence, Analysis, and Fate Using Mass Spectrometry*
Ángel Garciá and Yotis A. Senis (Editors) • *Platelet Proteomics: Principles, Analysis, and Applications*

Luigi Mondello • *Comprehensive Chromatography in Combination with Mass Spectrometry*
Jian Wang, James MacNeil, and Jack F. Kay • *Chemical Analysis of Antibiotic Residues in Food*