Springer Theses
Recognizing Outstanding Ph.D. Research

Casper Rønn Hoeck

Solving a 3D Structural Puzzle

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Casper Rønn Hoeck

Solving a 3D Structural Puzzle

Doctoral Thesis accepted
by the Technical University of Denmark,
Kongens Lyngby, Denmark

Springer

Author
Dr. Casper Rønn Hoeck
Department of Chemistry
Technical University of Denmark
Kongens Lyngby, Denmark

Supervisor
Prof. Charlotte H. Gotfredsen
Department of Chemistry
Technical University of Denmark
Kongens Lyngby, Denmark

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-319-96171-2 ISBN 978-3-319-96172-9 (eBook)
https://doi.org/10.1007/978-3-319-96172-9

Library of Congress Control Number: 2018947808

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Supervisor's Foreword

The establishment of molecular structure and configuration is central to our understanding of chemistry and biology—and many analytical techniques provide different structural knowledge in today's research. Two of the techniques most often employed for full structure elucidation are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Which analytical technique to use depends on the nature, origin, and properties of the compound being studied and the structural properties being researched.

NMR spectroscopy in structural studies of natural products and small molecules is an area of constant focus both in industry and academia where it is a very important part of the standard analytical toolbox in research and development. This requires a constant focus on developments within NMR and its applications, including the exploration of new approaches to employ NMR parameters in 3D structural calculations. The continued development of the entire structure elucidation process calls for a multidisciplinary approach, and the use of orthogonal experimental data is needed to increase the accuracy of the molecular structures.

The thesis by Casper Rønn Hoeck *Solving a 3D Structural Puzzle* describes different topics and novel research angles on how NMR spectroscopy can be employed in several interdisciplinary projects. The thesis describes development and implementation of new NMR pulse sequences, new computational approaches to structural calculations as well as organic polymer synthesis of alignment media—all of which contribute to obtaining more reliable NMR-based 3D molecular structures of small compounds.

In his work, Casper has contributed with significant new insights into NMR-based molecular structure determination and also highlighted some of the remaining challenges faced by 3D molecular structure elucidation of small compounds.

Kongens Lyngby, Denmark
March 2018

Prof. Charlotte H. Gotfredsen

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a versatile tool in analytical chemistry, highly suitable for structural elucidation of organic molecules—as well as multiple other areas of research. The subjects covered within this thesis all concern methods which allow a shift from covalent to spatial structural information using NMR spectroscopy.

Experimental distances from nuclear Overhauser effect (NOE) correlations, and dihedral angles from $^{3}J_{HH}$-coupling constants, were used to obtain 3D structural information for several natural and synthetic compounds. The stereochemistry of novel natural compounds was determined, including that of a bicyclic non-ribosomal peptide (a novel structural motif), a steroid, and several polyketides. Structural insights were gained for potential anti-cancer agents; the azumamides, including synthetic analogues. Differences in the conformational space of solution state compounds were identified experimentally between structural analogues and compared to the in vitro potency of the compounds. The structures of two peptides that exhibited a high degree of molecular recognition were investigated, resulting in the elucidation of a possible mode of interaction. Also, a major assumption in the calculation of distances from NOEs, the assumption of equal rotational correlation times between proton pairs, was investigated for molecules in organic solvents.

Two spin-state selective (S^3) HMBC experiments were developed for measurements of homonuclear and heteronuclear long-range coupling constants, respectively. The new NMR experiments were based on two existing experiments, the multiplicity-edited HMBC and the HAT HMBC, which were combined to obtain S^3 editing of long-range homonuclear coupling constants. The output of the first S^3 HMBC experiment was HMBC-type spectra with $^{n}J_{CH}$ correlated cross-peaks, from which $^{n+1}J_{HH}$-coupling constants were sign-selectively determined with high accuracy. Very small coupling constants, including previously unreported coupling constants from strychnine, were extracted, with all experimental values correlating very well to theoretical coupling constants from DFT calculation. A pulse segment was developed to change the polarization of the CH-H pairs in the homonuclear S^3 HMBC, to gain S^3 edited $^{n}J_{CH}$-coupling constants in the cross-peaks. While only determining coupling constants to methine carbons, the extracted experimental

coupling constants correlated very well to theoretical coupling constants, thus extending the S^3 HMBC methodology to include both $^{n+1}J_{HH}$- and $^{n}J_{CH}$-coupling constants.

Residual dipolar couplings (RDCs) are a relatively late addition to the small molecular NMR community, where alignment media are used to obtain anisotropic samples, which allow for RDCs to be extracted. The number of inter-nuclear vectors for the correlation of RDCs to 3D structures is often limited for small molecules. Homonuclear RDCs were extracted by use of the homonuclear S^3 HMBC that correlated well to alignment tensors from $^{1}D_{CH}$-coupling constants, thus increasing the number of internuclear vectors. The topic of enantiodiscrimination by RDC measurements of rigid organic molecules was also investigated, and new alignment media were developed to allow slight discrimination of enantiomers by stretched polymers. Finally, a new method of back-calculation of RDCs from 3D structures was developed and tested, which copes better with multiple conformers than the commonly used SVD methodology. The approach thus resulted in good conformer populations for several small molecules, including multiple cinchona alkaloids.

Preface

The work presented in this thesis was carried out during my three years of research as a Ph.D. student at the Department of Chemistry, Technical University of Denmark. The work was supervised by Associate Professor Charlotte Held Gotfredsen and funded by the Department of Chemistry. A three-month external stay was conducted in the group of Dr. Craig Butts at the School of Chemistry, University of Bristol, UK.

The thesis is divided into nine chapters, covering three main topics. The first topic is nuclear Overhauser effect correlations and $^{3}J_{HH}$-coupling constants in 3D structural investigations of organic compounds (Chap. 3). This is followed by a chapter concerning structural information from long-range coupling constants and new pulse sequences to determine these (Chap. 4). The final topic covers multiple chapters and focuses on 3D structural information from residual dipolar coupling constants, from increasing the number of available inter-nuclear vectors by long-range coupling constants (Chap. 6) to enantiodiscrimination (Chap. 7), and a chapter on structural flexibility and a new method of back-calculating (Chap. 8).

Kongens Lyngby, Denmark — Dr. Casper Rønn Hoeck

Parts of this thesis have been published in the following journal articles:

I. Maolanon, Alex R; Villadsen, Jesper S; Christensen, Niels J; **Hoeck, Casper**; Friis, Tina; Harris, Pernille; Gotfredsen, Charlotte H; Fristrup, Peter; Olsen, Christian A. Methyl Effect in Azumamides Provides Insight Into Histone Deacetylase Inhibition by Macrocycles. *Journal of Medicinal Chemistry*, **2014**, *57*, 9644–9657.

II. Petersen, Lene M; **Hoeck, Casper**; Frisvad, Jens C; Gotfredsen, Charlotte H; Larsen, Thomas O. Dereplication Guided Discovery of Secondary Metabolites of Mixed Biosynthetic Origin from *Aspergillus aculeatus*. *Molecules*, **2014**, *19*, 10898–10921.

III. Kjaerulff, Louise; Benie, Andrew J; **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. S^3 HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported J_{HH}. *Journal of Magnetic Resonance*, **2016**, *263*, 101–107.

IV. Li, Ming; **Hoeck, Casper**; Schoffelen, Sanne; Gotfredsen, Charlotte H; Meldal, Morten. Specific Electrostatic Molecular Recognition in Water. *Chemistry - A European Journal*, **2016**, *22*, 7206–7214.

V. **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. S^3 HMBC *hetero*: Spin-State-Selective HMBC for accurate measurement of heteronuclear coupling constants. *Journal of Magnetic Resonance*, **2017**, *275*, 68–72.

Acknowledgements

First and foremost, I would like to thank Lotte for accepting me as a Ph.D. student in her group, for giving me an opportunity to spend three years on exciting research within NMR spectroscopy, and for her guidance and support throughout. It has been a privilege, and I have learned a lot in the three years. I would like to thank the entire NMR group, especially to former group member Louise Kjærulff for great NMR talks, discussions, and cooperation on new NMR experiments and Anne Hector for always being very helpful. Regarding NMR experiments, I would also like to thank Ole W. Sørensen for invaluable help on theoretical NMR spectroscopy in our collaborations. Many thanks to Sebastian Meier for always being helpful and especially for proofreading my thesis. For this, I also owe Niclas Hoeck and Pernille Rønn-Nielsen my thanks. Thanks go to the people of building 201 (now 211) and the rest of the Department of Chemistry, for three highly enjoyable and memorable years.

I am thankful that I through my studies have had extraordinary collaborators: Christian A. Olsen, KU, and former group members Alex Maolanon and Jesper Villadsen with whom I was involved in work regarding the 3D structures of azumamides and Morten Meldal, KU, and Ming Li with whom we collaborated on projects regarding NMR investigations of molecular recognition of synthetic peptides. I would also like to thank Thomas Ostenfeld Larsen, Bioengineering, DTU, and Thomas Isbrandt for our continued collaboration on secondary metabolites from *A. homomorphus* as well as the former group member Lene Pedersen for shared work on 3D structures of metabolites from other fungi.

I am grateful to Craig for hosting me as a visiting scholar and the entire NMR group (Craig, Jess, Iyke, Claire, Tom, Siying, and Paul) at the School of Chemistry, University of Bristol, for being very welcoming and making my three-month external stay enjoyable and rewarding.

Last but not least, my deepest thanks to my family, my friends, and Pernille, for unwavering encouragement and support.

Kongens Lyngby, Denmark
May 2016

Dr. Casper Rønn Hoeck

Contents

Abbreviations

1D	One-dimensional
2D	Two-dimensional
2F	Two-field
3D	Three-dimensional
AIBN	Azobisisobutyronitrile
AP	Anti-phase
APhES	2-acrylamido-2-phenylethane-1-sulfonic acid
AU	Automation
ax	Axial
Azu A	Azumamide A
B3LYP	Becke's three-parameter hybrid functionals
C10-Phe	Decyl-acryloyl-L-phenylalaninate
$CHCl_3$	Chloroform
CM-A	Cyclomorphosin A
CM-B	Cyclomorphosin B
CORMA	Complete relaxation matrix analysis
COSY	Correlation spectroscopy
CSGT	Continuous set of gauge transformation
CW	Continuous wave
DCM	Dichloromethane
Des C	Desmethyl-azumamide C
DFT	Density functional theory
DMAA	(*N,N*)-dimethyl acrylamide
DMSO	Dimethyl sulphoxide
DQ	Double quantum
DQF	Double quantum filtered
DSO	Diamagnetic spin–orbit
DTU	Technical University of Denmark
EGDMA	Ethylene glycol dimethylacrylate
Epi E	Epi-azumamide E

eq	Equatorial
ESI	Electrospray ionization
EtOAc	Ethyl acetate
F1	Indirect dimension
F2	Direct dimension
FC	Fermi contact
FDA	Food and drug administration
FDAA	1-Fluoro-2,4-dinitrophenyl-5-L-alanineamide
FF	Force field
FID	Free induction decay
FT	Fourier transform
GCB	Generalized cosine β
GIAO	Gauge-invariant atomic orbitals
GTO	Gaussian-type orbital
HAT	Homonuclear *J*-attenuated
HDAC	Histone deacetylase
HLA	Haasnoot, De Leeuw and Altona (or Altona)
HMBC	Heteronuclear multiple bond correlation
HSQBC	Heteronuclear single quantum multiple bond connectivity
HSQC	Heteronuclear single quantum coherence
INEPT	Insensitive nuclei enhanced by polarization transfer
IP	In-phase
IPAP	In-phase/anti-phase
IPC	Isopinocampheol
IRA	Initial rate approximation
ISPA	Isolated spin-pair approximation
KU	University of Copenhagen
L7	Ligand 7
LC	Liquid chromatography
LC	Liquid crystal
MA	Mass spectrometry
MAE	Mean absolute error
MARDIGRAS	Matrix analysis of relaxation for discerning the geometry of an aqueous structure
MCMT	Multi-conformer–multi-tensor
MCST	Multi-conformer–single-tensor
MD	Molecular dynamics
MeOH	Methanol
MM	Molecular mechanics
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOE/ROE	NOESY/ROESY comparison
NOESY	Nuclear Overhauser effect spectroscopy
NRP	Non-ribosomal peptide
PANIC	Peak amplitude normalization for improved cross-relaxation

PBLG	Poly-γ-benzyl-L-glutamate
PDMAA	Poly-(*N,N*)-dimethyl acrylamide (crosslinked)
PEGA	Poly(acryloyl-bis-(aminopropyl)polyethylene glycol)
PhEtN	1-Phenyl ethylamine
PhEtO	1-Phenyl ethanol
ppm	Parts per million
PSO	Paramagnetic spin–orbit
Q	Q factor
RDC	Residual dipolar coupling constant
RF	Radio frequency
RMSD	Real mean square deviation
ROESY	Rotating frame nuclear Overhauser effect spectroscopy
S/N	Signal-to-noise
S^3	Spin-state selective
SAG	Strain-induced alignment in a gel
SD	Spin-dipole operator
SM	Secondary metabolite
SO	Spin–orbit coupling
SP	Stretched polymer
SQ	Single quantum
STO	Slater-type orbital
SVD	Singular value decomposition
T_1	Longitudinal relaxation
T_2	Transverse relaxation
T20	Target 20
TEA	Triethyl amine
TEG-Phe	(Triethylene glycol methyl ether)-acryloyl-L-phenylalaninate
TMS	Tetramethylsilane
TOCSY	Total correlation spectroscopy
UHPLC	Ultrahigh-pressure liquid spectroscopy
UV	Ultraviolet
YES	Yeast extract sucrose agar
ZQ	Zero-quantum
B_0	Magnetic field
ω	Frequency in rad
γ	Gyromagnetic ratio
$\hbar$	Reduced Planck constant
μ_0	Vacuum permeability constant
σ	Cross-relaxation rate constant
ρ	Dipolar longitudinal relaxation rate constant
η_{IS}	Cross-peak intensity
η_I or η_S	Diagonal/irradiated peak intensity
τ_c	Rotational correlation time
τ_s	Relative intramolecular movements
τ_o	Overall rotational correlation time

τ_m	Mixing time
J	J-coupling constant
D	Dipolar coupling constant
T	Total coupling constant
A_a,D_a	Axial component, degree of alignment
A_r,D_r	Rhombic component
R	Rhombicity
D_{max}	Maximum possible dipolar coupling constant
$D_{exp,max}$	Maximum experimental RDC
$D_{exp,min}$	Minimum experimental RDC

List of Figures

List of Tables

Chapter 1
NMR Spectroscopy: Past to Present

The foundation for modern day NMR spectroscopy was laid more than 70 years ago, in 1946 by the work of the groups of Bloch and Purcell who first reported on NMR in bulk materials. Work that awarded them a shared Nobel Prize in Physics in 1952 [1, 2]. One could look back even further to the introduction of the concept of nuclear spin, the phenomenon exploited in NMR spectroscopy, which was described in the 1920s [3, 4]. Since the early discoveries much has happened, from the introduction of superconducting magnets and development of high field instruments, to proceedings in methods of data acquisition, leading to today where NMR spectroscopy is one of the most widely used spectroscopic techniques in chemistry, structural biology and material sciences—and arguably the most information rich when used appropriately. It is based on an electron density dependent energy difference of nuclear spin orientation in an applied magnetic field, as outlined in Eq. (1.1), and the ability of interacting with these spins using radio frequency (RF) energy.

$$\omega = \gamma B_0(1 - \sigma) \tag{1.1}$$

where ω is the resonance frequency of a given nucleus, γ is the gyromagnetic ratio of said nucleus, B_0 is the applied magnetic field and σ is the shielding constant which is dependent on the electron density and configuration of electrons around the nucleus [3].

In the earliest days of NMR spectroscopy the technique used was continuous-wave spectroscopy, where the frequency of the RF pulse was varied, and the spectra recorded by scanning the spectral range from start to finish [3, 4]. This changed in the mid-1960s with the introduction of Fourier transform (FT) techniques by Ernst and Anderson; work for which Ernst was rewarded with the 1992 Nobel Prize in Chemistry [3, 5]. FT is a method to extract frequencies from free induction decay (FID), and thus data for the full spectral range may be acquired from a single short RF pulse. NMR spectroscopy has since evolved into a versatile field, largely in part of ever evolving methods of using RF pulses to

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_1

influence targeted nuclei and output a diverse range of spectra, each with a specific purpose. The large array of pulse sequences, which each holds the promise of a vast amount of information, means that NMR is applicable when investigating very different systems; small molecules, macromolecules and materials in liquid- or solid state alike.

On the hardware side, the introduction of superconducting magnets in the mid-1960s laid the ground for modern day spectrometers [3, 4]. Until then 100 MHz was the maximum field strength obtainable, but the development of higher field instruments were available through the new technology, from 220 MHz in 1966, 500 MHz in 1978, 800 MHz in 1995 to modern day instruments pushing the 1 GHz boundaries and beyond. Due to the correlation of magnetic field strength, resolution and signal-to-noise (*S/N*), the shift to bigger magnets has vastly increased the scope of the technology, hand in hand with the development of better probes and other relevant hardware.

This thesis will primarily focus on unique possibilities of NMR spectroscopy that sets the method apart from many other spectroscopic methods, namely observable interactions between nuclei. These magnetic dipole interactions are used extensively in this thesis and are the focus of the theory section.

References

1. E.M. Purcell, H.C. Torrey, R.V. Pound, Phys. Rev. **69**, 37–38 (1946)
2. F. Bloch, W.W. Hansen, M. Packard, Phys. Rev. **69**, 127 (1946)
3. E.D. Becker, Anal. Chem. **65**, 295A–302A (1993)
4. D.L. Rabenstein, Anal. Chem. **73**, 214A–223A (2001)
5. R.R. Ernst, W.A. Anderson, Rev. Sci. Instrum. **37**, 93–102 (1966)

Chapter 2
Theory—3D Structural Information from NMR, Part 1

The various projects included in this thesis share a common trait—they all rely on information obtained from magnetic dipolar interactions between nuclei in organic molecules.

The main focus will be on NMR observables which are useful in elucidations and calculations of 3D structures, namely nuclear Overhauser effect correlations (NOEs) (dipolar relaxation, through space), *J*-coupling constants (indirect dipole-dipole couplings, through bonds) and residual dipolar coupling constants (RDCs) (direct dipole-dipole couplings, through space). The three subjects are conceptually depicted in Fig. 2.1 and each topic will be treated independently in this chapter and Chap. 5.

Chosen theory for the different topics is presented, and focus will be on the theory which was applied, either directly or implicitly. This selection results in some rather broad jumps in the presented theory and further information may be found in more comprehensive literature sources on the subjects, e.g. from the cited references.

2.1 The Nuclear Overhauser Effect

The mechanism involved in nuclear Overhauser effect (NOE) spectroscopy is dipolar cross-relaxation of nuclei in close spatial proximity. Being facilitated entirely through space in contrast to e.g. the bond mediated *J*-coupling constants, *vide infra*, NOEs lead to purely spatial information of the nuclei, as illustrated in Fig. 2.2 [1, 2].

The phenomenon is named after Overhauser, who first proposed that polarization of metal nuclear spins would be observable upon saturation of the metal electrons in 1953 [3, 4]. Solomon established the theoretical basis and first experimental evidence of NOEs two years later. The proposed Solomon equation, Eq. (2.1), is thus considered the basis of NOE theory [2, 5]. While the NOE effect is equally

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_2

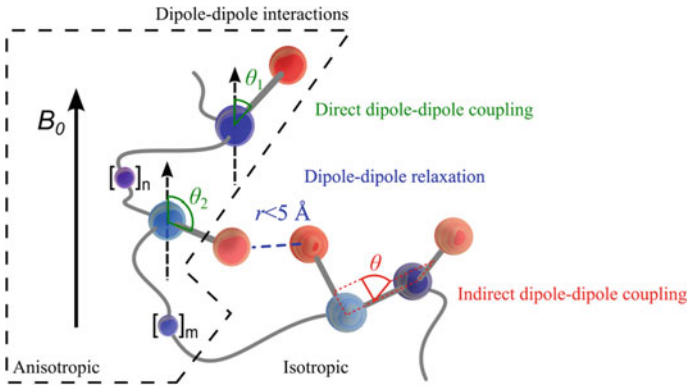


Fig. 2.1 Concepts of magnetic dipole-dipole interactions and the obtainable 3D structural information. In green are the direct dipole-dipole couplings, which may be used to gain information of angles and distances of internuclear vectors independent on their situation in the molecular structure. In blue is the dipole-dipole relaxation, which may give access to local internuclear distances through space. Indirect dipole-dipole couplings are highlighted in red which, among multiple uses, hold local information of the relative position of nuclei, mediated through the bonds in the structure. It is indicated whether the information may be extracted from spectra of molecules under isotropic conditions or if anisotropic conditions are also needed

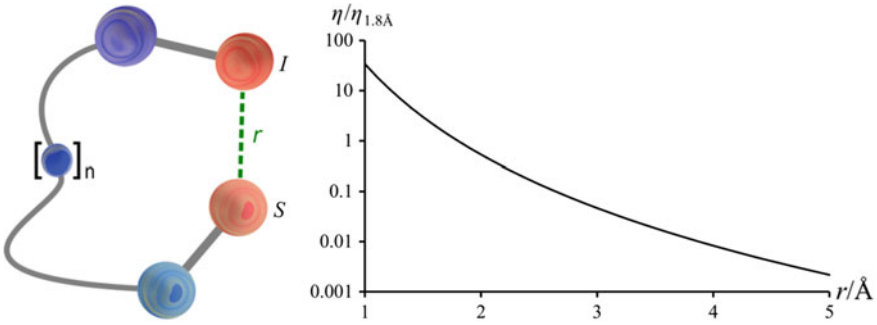


Fig. 2.2 Left: Illustration of the through-space dependance of NOE relaxation of protons (red), blue may be carbon or heteroatoms. Right: Correlation of theoretical relative NOE intensity and distance (in Å) between protons. Relative to a reference distance of 1.8 Å set to 1. Note the logarithmic scale of the y-axis

important in homo- and heteronuclear NMR, the latter exemplified from e.g. NOE enhancement of ^{13}C resonances, the homonuclear variant between protons is a prime source of structural information and is the focus of this section. Note that the involved nuclei will be labeled as I and S, even though both are protons (in some publications I_1 and I_2 or other notation is used for homonuclear cases).

2.1.1 Relaxation

The generation of NOE correlations is a relaxation process and some points regarding general relaxation processes in liquid NMR spectroscopy will be made, while the literature provides a much more in depth description [2]. Relaxation in NMR is the mechanism of returning to equilibrium after the spin population has been perturbed by RF pulses and is divided into longitudinal (T_1) and transverse (T_2) relaxation. In short T_1 relaxation is the recovery of the original longitudinal magnetization, and T_2 is the loss of transverse magnetization [2]. For small organic molecules, which are the focus of this thesis, the timeframes of T_2 and T_1 relaxation processes are usually of equal orders of magnitude, while T_2 may be much shorter than T_1 for macromolecules or solids [6].

The longitudinal relaxation is the main interest for understanding NOE relaxation. The relaxation process is not caused by spontaneous emission, as this is slow for the energy-levels involved in NMR spectroscopy. Instead it is facilitated by the translations, rotation and internal motions of the molecules in the sample—termed "the lattice" [2]. This may describe the difference between the relaxation properties of different sizes of molecules, as the motions, or tumbling, will be quite different. The energy from the relaxation is taken up by the lattice as the near infinite degrees of freedom of the full lattice translate to the energy levels of the lattice being a continuum. This means that any NMR transition will have a matched energy level in the lattice.

The relaxation mechanism of the NOE relaxation is dominated by the contribution from the relative spin dipole reorientation of neighboring nuclei under rapid isotropic tumbling, as the population of one spin changes when the population of another nuclear spin is perturbed [1, 2]. This is the cause of the dependence of NOEs on the rotational correlation time (τ_c). The initial population shift of a nuclear spin can occur by continuous saturation of a resonance, as illustrated in Fig. 2.3, the basis of steady-state NOEs. The continuous saturation effectively leads to equal nuclear spin population differences across the corresponding α/β transitions. The NOE is given by the W_2 and W_0 cross-relaxation where both spins switch spin state simultaneously, see the legend of Fig. 2.3 for more information.

The steady state situation in Fig. 2.3 is described by the Solomon Eq. (2.1), which defines the change of I_z (a vector equal to the population difference between the spin states) over time as a function of the relaxation pathways and the initial population differences I_z^0/S_z^0. $\partial I_z/\mathrm{dt}$ and S_z are equal to 0 for steady state [2].

$$\frac{\partial I_z}{\partial t} = -\left(I_z - I_z^0\right)\left(W_{0IS} + 2W_{1I} + W_{2IS}\right) - \left(S_z - S_z^0\right)\left(W_{2IS} - W_{0IS}\right) \tag{2.1}$$

This equation may be rewritten to describe the NOE enhancement, $f_I\{S\}$, which describes the change of intensity for spin I when saturating spin S, using that $S_z^0 = (\gamma_S/\gamma_I)I_z^0$ [2].

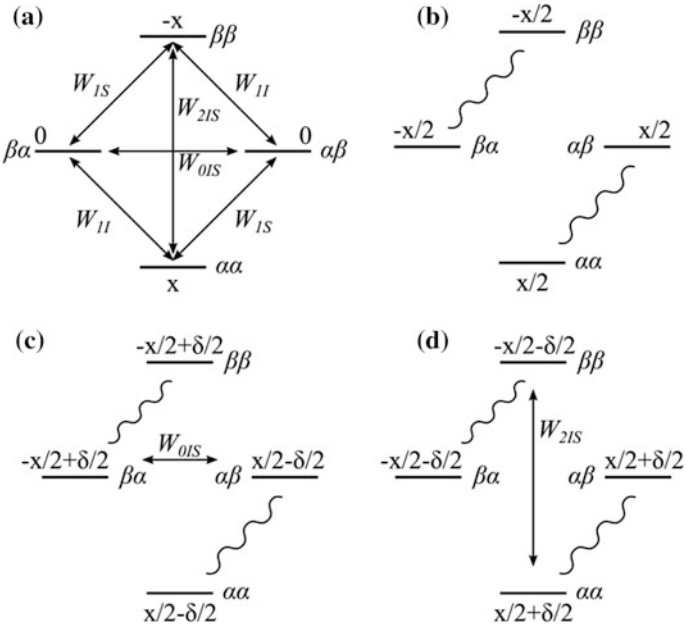


Fig. 2.3 Energy diagrams for the NOE in a homonuclear two-spin system of the protons I and S. The intensity of proton I is proportional to $(N_{\alpha\alpha} - N_{\beta\alpha}) + (N_{\alpha\beta} - N_{\beta\beta})$. Inspired by Neuhaus and Williamson [2]. **a** The possible spins states, given as $I(\alpha)S(\beta)$, and transition probabilities (W). The W_2 transition (the double quantum) and the W_0 (zero quantum) transition are the relevant cross-relaxation pathways. The initial populations are $N_{\alpha\alpha} = x$, $N_{\beta\beta} = -x$ and $N_{\alpha\beta}/N_{\beta\alpha} = 0$. **b** Saturation of the S resonance leads to $N_{\alpha\alpha} = N_{\alpha\beta}$ and $N_{\beta\beta} = N_{\beta\alpha}$. The saturation is maintained throughout. **c** The effect of W_0 cross-relaxation leads to a transfer of the population δ/2 from $N_{\alpha\alpha}$ to $N_{\beta\beta}$. This leads to a shift towards equilibrium and thus a decrease in the intensity of resonance I; a negative NOE enhancement. **d** The effect of W_2 cross-relaxation leads to a transfer of the population δ/2 from $N_{\beta\beta}$ to $N_{\alpha\alpha}$. This leads to a shift away from equilibrium and thus an increase in the intensity of resonance I; a positive NOE enhancement

$$f_I\{S\} = \frac{\left(I_z - I_z^0\right)}{I_z^0} = \frac{\gamma_S}{\gamma_I} \frac{\left(W_{2IS} - W_{0IS}\right)}{\left(W_{0IS} + 2W_{1I} + W_{2IS}\right)} = \frac{\sigma_{IS}}{\rho_{IS}} \tag{2.2}$$

where σ is the cross-relaxation and ρ is the dipolar longitudinal relaxation rate constant. The cross-relaxation rate determines the intensity of the cross-peaks and is of major interest. It is positive if W_{2IS} dominates W_{0IS}, and negative in the opposite scenario, which is in correlation to what may be found from Fig. 2.3. The transition probabilities, or spectral densities, may be calculated from the dipolar interaction Hamiltonian to give the formula for cross-relaxation shown in Eq. (2.3) [2], where it is utilized that the transition probabilities are proportional to the chemical shift frequencies of the nuclei as the $\alpha\beta \leftrightarrow \beta\alpha$ transition corresponds to a frequency of $|(\omega_I - \omega_S)|$ and the $\alpha\alpha \leftrightarrow \beta\beta$ transition corresponds to a frequency of $(\omega_I + \omega_S)$, which may also be realized from Fig. 2.3.

$$\begin{aligned}\sigma_{IS} &= W_{2IS} - W_{0IS} = \frac{1}{10}K^2\tau_c\left(\frac{6}{1+(\omega_I+\omega_S)^2\tau_c^2} - \frac{1}{1+(|\omega_I-\omega_S|)^2\tau_c^2}\right) \\ &= \frac{1}{10}K^2\tau_c\left(\frac{6}{1+(\omega_I+\omega_S)^2\tau_c^2} - 1\right) \\ K &= \left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_I\gamma_S r^{-3}\end{aligned} \tag{2.3}$$

where γ is the gyromagnetic ratio of the nuclei and r is the distance between the involved nuclei, μ_0 is the vacuum permeability constant and $\hbar$ the reduced Planck constant. Equation (2.3) exploits that the difference between ω_I and ω_S is much lower than one, especially in the narrow chemical shift range of protons. In the same manner the dipolar longitudinal relaxation rate may be calculated [2].

All of these equations are based on the steady state experiment which is seldom used today, but more easily explained. The more utilized experiments to obtain NOE spectra is based on inversion, where nuclear spin-population differences are inverted across the transitions, which leads to kinetic NOEs. The method is the most frequently utilized due to fewer artifacts [2]. The kinetic NOE is used to determine the rate of which the steady state is reached instead of performing measurements on the steady state itself, and is implemented as either the truncated NOE (TOE) or as the transient NOE [2]. In transient NOE experiments, which are used throughout this thesis, the resonances are initially inverted and the NOE evolves without RF irradiation for a fixed period, termed the mixing time τ_m. After the evolution an observe pulse is applied to sample the population distribution and obtain the NOE spectra. In this type of NOESY experiment the NOE enhancements initially build up linearly, reach a maximum where linearity is lost and then decay back to zero.

2.1.2 The ROESY Experiment

The rotating frame NOE (ROESY) experiment, initially dubbed CAMELSPIN (cross-relaxation appropriate for mini-molecules emulated by spin-locking) is a related experiment to the NOESY experiments which evolves the enhancements between elements of transverse magnetization instead of longitudinal magnetization [2, 7–9]. For this purpose, a spin-lock from continuous wave (CW) or a pulse train of multiple 180° pulses (T-ROESY) is used for a fixed period of time. The spin-lock period is similar to the mixing time of the transient NOESY experiments. This is advantageous for small molecular NMR, as the ROESY variant of the NOESY longitudinal cross-relaxation depends differently on the rotational correlation time as seen in given in Eq. (2.4) for CW-ROESY and (2.5) for T-ROESY [2, 7, 8, 10].

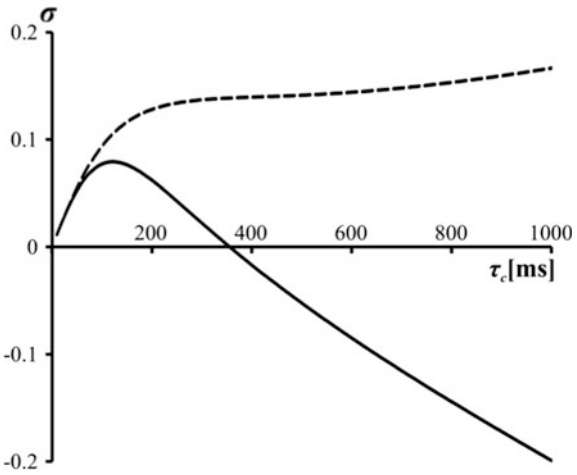


Fig. 2.4 The theoretical cross-relaxation rate (σ) of NOE (—) and T-ROE (- -) experiments at 500 MHz, for a distance of 2.5 Å. Note that the crossover point is dependent on the field strength. The cross-relaxation rate is shown in contrast to the often depictured maximum homonuclear enhancement, as the cross-relaxation rate is the actual observable parameter in NOE experiments. Calculated using Eqs. (2.3) and (2.5)

$$\sigma_{IS} = \frac{1}{10} K^2 \tau_c \left(\frac{3}{1+\omega^2\tau_c^2} + 2 \right) \tag{2.4}$$

$$\sigma_{IS} = \frac{1}{20} K^2 \tau_c \left(\frac{6}{1+4\omega^2\tau_c^2} + \frac{3}{1+\omega^2\tau_c^2} + 1 \right) \tag{2.5}$$

For small and medium size molecules at a certain rotational correlation time, the NOE enhancement is zero, which is avoided in the ROESY experiments as seen in Fig. 2.4 for NOE and T-ROE.

While there are clear cases where the utilization of ROESY experiments are advantageous due to a better signal to noise ratio, ROESY spectra suffer from TOCSY transfers and a offset-dependency of the spin-lock, which needs to be accounted for [2, 7]. This makes the ROESY data harder to analyze and NOESY spectra were generally used where possible in this thesis. For more on NOE/ROE theory, related to τ_c, see Sect. 3.5.

2.1.3 NOE in Structure Determination

The NOE correlations are usually used qualitatively to determine stereochemistry or confirm the assignment of nuclei in novel or known compounds [2, 11]. The information is highly local due to a $1/r^6$ dependence on the intensity of the observed signal and distances up to 4–5 Å are usually observed, depending on the compound size in the form of the rotational correlation time of the nuclei pair.

The rotational correlation time is dependent on the size of the molecule, the temperature, the shape and flexibility of a structure and the viscosity of the solvent. The signal from NOE spectroscopy may be quantified if care is taken in the experimental setup and some approximations are made, presented below.

In general, two methods are available for obtaining quantitative distances from NOE data; with or without the use of a relaxation matrix [12–16]. The difference between these methods is in short the approach to handling multi-spin effects such as unwanted spin diffusion. Spin diffusion may be explained from a situation where three spins (H_a, H_b and H_c) are considered, as in Fig. 2.5.

Spin diffusion will create a cross-signal between H_a and H_c with a stronger intensity than expected from the inter-proton distance, and the signal may lead to a false stereochemistry or experimental distance. This problem may be avoided or diminished by using short mixing times as magnetization has less time to build on H_b in Fig. 2.5 [2, 17]. When determining whether to use a relaxation matrix method or not, it is important to consider the system; for large molecules where spin diffusion is rapid, it should be included in the calculations. For small molecules, using short mixing times, spin diffusion is less of a problem. Thus, using a matrix that also includes a multitude of approximations may not be needed or beneficial.

Another important consideration in the experimental setup is the utilized delay time between scans, which is true for all quantitative NMR. If inadequate T_1 relaxation is achieved for any protons during the delay, the intensity of subsequent scans will be altered accordingly leading to wrong relative intensities and thus wrong distances, see below.

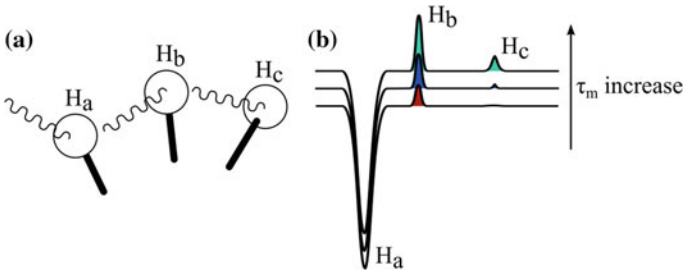


Fig. 2.5 Spin diffusion explained. **a** Three spins are considered (H_a, H_b and H_c). H_b is positioned between H_a and H_c, and is involved in NOE relaxation with both, while H_a and H_c are far from each other with negligible cross-relaxation. Magnetization is transferred from H_a to H_b, and once sufficient magnetization has built up on H_b, magnetization transfers (or diffuses) from spin H_b to H_c. **b** An illustration of a 1D- or slice of a 2D NOESY spectrum shows the NOE cross-peak intensity increasing with an increase in the used mixing times

2.1.4 Quantitative Calculations

The nuclear Overhauser effect (NOE) gives through-space correlations from dipolar relaxation between nuclei. The relationship between cross-relaxation, σ, and distance, r, is given in Eq. (2.3) and rewritten below to match most litterature [2].

$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2 \gamma_I^2 \gamma_S^2}{10} r^{-6} \left(\frac{6\tau_c}{1+4\omega^2\tau_c^2} - \tau_c\right) \tag{2.3}$$

This is easily simplified, assuming identical nuclei in all experiments, to Eq. (2.6), where k covers the constants in (2.3) and is identical for all observed NOEs.

$$\sigma_{IS} = kr^{-6} \left(\frac{6\tau_c}{1+4\omega^2\tau_c^2} - \tau_c\right) \tag{2.6}$$

Assuming identical correlation time for all nuclei pairs, it is further simplified to Eq. (2.7). Again k covers constants and is equal for all NOEs.

$$\sigma_{IS} = kr^{-6} \tag{2.7}$$

The cross-relaxation rate for transient NOEs is correlated to the observed intensity, η, by Eq. (2.8).

$$\eta_{IS} = \sigma_{IS}\tau_m \tag{2.8}$$

where τ_m is the mixing time used in the experiment. If identical mixing times are used, the relationship of intensities and distances between two nuclei pairs may be defined as in Eq. (2.9).

$$\frac{\eta_1}{\eta_2} = \frac{kr_1^{-6}}{kr_2^{-6}} \rightarrow r_1 = r_2 \left(\frac{\eta_1}{\eta_2}\right)^{-\frac{1}{6}} \tag{2.9}$$

Thus, if a single distance is known, others may be calculated from their relative intensity. This relationship is known either as the isolated spin-pair approximation (ISPA) or initial rate approximation (IRA) [17, 18]. As indicated this is an approximation due to a couple of factors.

1. It is approximated that all rotational correlation times are equal, and unknown dynamics thus affect the NOE.
2. It is assumed that the acquired mixing time is short, eliminating any effect from spin diffusion.
3. The cross-relaxations of multiple involved spins are assumed to behave as isolated spins.

In order to avoid the inclusion of spin diffusion contribution to the NOE intensities in calculations, the mixing time may be varied to construct build-up curves to ensure that the utilized mixing time is in the linear regime, exemplified in Fig. 2.6. For spin diffusion the onset of magnetization buildup will be later than for "true" NOE cross-peaks making them easily distinguishable. This also allows for the elimination of possible noise or artifact signals in the subsequent analysis, as these signals will be completely independent of the mixing time.

There are two methods to proceed once the linear range has been established. One is to use the cross-relaxation rate which may be determined directly from the lower graph of Fig. 2.6 (right) as the intersection with the y-axis. The other is to utilize a fixed mixing time in the linear range. While the first method will effectively average out some of the internal intensity errors in the spectra, the latter is usually used in this thesis, due to interferences from COSY-type artifacts in the spectra, especially at low mixing times due to the lower NOE intensities [2]. To use points within the linear range, where the scalar coupling contribution is negligible, leads to the inclusion of more NOEs, and thus more information, and the associated errors were always evaluated to be within the general error of the experiments. It is here utilized that the scalar coupling intensities are less dependent on the mixing time than the NOE cross-peak intensities.

In theory, any distance may be used as reference distance, as all distances are used relative to each other. In practice, the most common distances chosen are those known to be static; diastereotopic protons ($\sim$1.78 Å) or neighboring aromatic protons ($\sim$2.48 Å). The signal intensity of both of these may be modulated due to scalar couplings, which should be kept in mind. Furthermore, a dependence on the

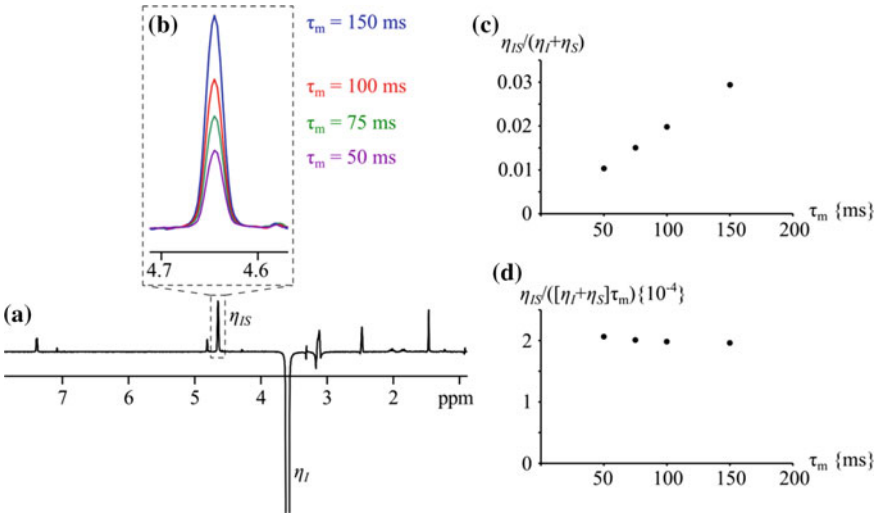


Fig. 2.6 **a** 1D slide from 2D NOESY spectrum of epi-10,23-dihydro-24,25-dehydroaflavine (see Sect. 3.2.2). The dependence of the intensity on the mixing time (τ_m) is seen in the enlargement (**b**) and is plotted on the right in a buildup curve for the exemplified NOE (**c** and **d**)

value of the chosen reference distance and the error distribution has been reported, as the distances of the same size resulted in a better fit between experimental and back-calculated data [2]. That being said, excellent fits have been reported utilizing e.g. diastereotopic protons, which are on the low side of relative distances in organic structures [16]. Note that the set distance of the reference may be allowed to differ (e.g. up to 5% error), to obtain the best possible correlation between experimental and 3D structure distances. This is most easily achieved on rigid systems, with multiple fixed distances.

2.1.5 The 1D NOESY/ROESY (PANIC) Approach

The PANIC or Peak Amplitude Normalization for Improved Cross-relaxation method is, other than a testament of the creative acronyms created by the NMR community, a straight forward implementation of the ISPA approach [18]. It is based on the work of Macura et al. discussed in the next section, but is used primarily for 1D NOESY/ROESY data [12]. It is used that the linear range of the NOE buildup curve is quite remarkably extended by implementing the auto-relaxation of the irradiated nuclei into the measurement of the NOE intensities in Eq. (2.10) [12, 18].

$$\eta_{IS}^{PANIC} = \frac{\eta_{IS}}{\eta_I} \tag{2.10}$$

In practice this is easily achieved by setting the integral of the irradiated peak to be equal among all 1D spectra obtained and these "PANIC intensities" are compared in the ISPA method. The resulting relative intensities are thus corrected for auto-relaxation. The irradiated peak is set to a value of −1000, chosen for convenience as this makes the relative magnitude of the cross-peaks in the order of 0–500 for small molecules, dependent on τ_c and τ_m.

2.1.6 The 2D NOESY/ROESY Approach

As mentioned in the previous section, the PANIC approach is based on the work of Macura et al. for 2D spectra [12, 18]. Thus, when dealing with 2D NOESY or ROESY spectra the methodology outlined here is used. It is essentially equal to the PANIC approach, but instead of using a single irradiated peak as reference, the cross-peaks (η_{IS}) are normalized by the average of the two appropriate diagonal peaks (η_I and η_S) as in Eq. (2.11).

$$\eta_{IS}^{norm} = \frac{\eta_{IS}}{1/2(\eta_I + \eta_S)} \tag{2.11}$$

If a lack of resolution leads to overlap in the diagonal peaks, Eq. (2.12) may be used, as it is assumed that η_I and η_S are close to identical for equal mixing time and equal nuclei.

$$\eta_{IS}^{norm} = \frac{\eta_{IS}}{\eta_I} = \frac{\eta_{IS}}{\eta_S} \tag{2.12}$$

In practice it may be beneficial to divide overlapping diagonal peaks by the number of protons present, as the auto-relaxation is more likely to be equal for an identical electronic environment and use this average diagonal peak in the calculations [19].

2.1.7 Relaxation Matrix Approach

The relaxation matrix approach was not generally utilized in the work described in this thesis, though it was evaluated initially. A short introduction will thus suffice. The Solomon equation, which is given in Eq. (2.1), may be re-written in matrix form where cross-peak intensities are considered as a function of mixing time, τ_m [2, 17, 20, 21].

$$\frac{\partial A(\tau_m)}{\partial \tau_m} = -RA(\tau_m) \tag{2.13}$$

$$A(\tau_m) = A_0 e^{-R\tau_m} \tag{2.14}$$

where $A(\tau_m)$ is the matrix of cross-peak intensities at τ_m, and R is the symmetrical relaxation rate matrix given in Eq. (2.15). A_0 is a diagonal matrix consisting of the cross-peak intensities at $\tau_m = 0$ s.

$$R = \begin{bmatrix} \rho_1 & \sigma_{12} & \cdots \\ \sigma_{21} & \rho_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{2.15}$$

The diagonal elements of the relaxation matrix are the dipolar longitudinal relaxation rate constants, ρ, and the off-diagonal elements are the cross-relaxation rate constants, σ. From these the inter-proton distance between nuclear spins may be obtained.

This may be re-arranged to (2.16) to express the relaxation rates in terms of NOE intensities.

$$-\ln\left(\frac{A}{A_0}\right)\tau_m = R \tag{2.16}$$

Which may be expressed and solved as (2.17), where the matrix E contains the eigenvector matrix and N contains the eigenvalues of A/A_0.

$$\frac{-E \ln NE^{-1}}{\tau_m} = R \tag{2.17}$$

Cross-relaxation rates can thus be derived and used to obtain inter-nuclear distances. The NOE intensities may also be back-calculated from a molecular 3D structure using the inverse of this method to generate a theoretical NOE spectrum, which can be iteratively compared to the experimental NOE spectrum for structure refinement and verification [17, 20, 21].

If the matrix works perfectly in its implementation the method takes multi-spin effects, such as spin diffusion, into account. It is thus most relevant when dealing with macromolecular systems where multi-spin effects are generally more prevalent. Problems can occur for full relaxation matrix analysis where incomplete data sets (e.g. due to spectral overlap or spectral noise) make derivation of the full intensity matrix difficult. This is circumvented by combining experimental data with cross-peak intensities calculated from a theoretical model by programs that utilize the full relaxation matrix method such as CORMA (Complete Relaxation Matrix Analysis) and MARDIGRAS (Matrix Analysis of Relaxation for discerning the Geometry of an Aqueous Structure) [17, 21].

2.1.8 Averaging of NOE Data

When an average of different nuclei is observed, e.g. from overlapping resonances of methylene- or methyl groups, averaging methods are needed in order to cope. In non-methyl cases, Eq. (2.18) is used in distance calculations.

$$\eta_{I(Sa+Sb)} = k\left(r_{ISa}^{-6} + r_{ISb}^{-6}\right) \tag{2.18}$$

For methyl groups another approach is taken. The traditional methods to average methyl groups are derived for protein NMR and are split in two, dependent of the relative local rotational correlation time of the methyl group compared to the rest of the structure, given in Eqs. (2.19) and (2.20) [22].

$$\eta_{IMe} = k\left\langle r_{IMe}^{-3}\right\rangle^2 \tag{2.19}$$

$$\eta_{IMe} = k\left\langle r_{IMe}^{-6}\right\rangle \tag{2.20}$$

If the methyl group is spinning much faster than the rotational correlation time for the general molecule, the NOE is averaged as (2.19) and if the effective rotational correlation times are close to equal (2.20) is used [22]. In practice (2.20) is only used for very small molecules in very low viscosity solvents.

2.1.9 Handling of Rigid and Flexible Molecules

NOE distance analysis has proven very accurate for small molecules when few conformers are present [16, 23–25]. From NOE distances another low abundance conformer was thus shown to be present for strychnine [25]. For more flexible compounds it is beneficial to rely on less optimized structures or on centroids which represent the full conformational space. The usage of centroids is a laborious task though; one may need to use many centroids if the compound is very flexible. One also has to identify precisely what parts are of interest for the structure as the nuclei used in establishing the centroids. The inclusion of multiple structures requires means of averaging a vast amount of structures in a sensible way. Overfitting or fitting of data to a wrong conformational average may here present a problem since multiple different conformational populations may correlate well to the distances obtained from the experimentally observed NOEs [26, 27]. It is thus recommended to couple the NOE data to other types of NMR data, being either *J*-coupling or residual dipolar coupling, see Sects. 2.2 and 5.1. This will alleviate some problems due to the difference in the averaging mechanisms for the different data; averaging distances, dihedral angle, angles etc.

When comparing multiple structures to NOE derived distances, the average distances are calculated by equation

$$r_{ave} = \left\langle p_i r_i^{-6} \right\rangle^{-1/6} \tag{2.21}$$

where p_i is the population with a given distance. The averaging results in a larger influence of shorter distances to the observed average distance [25].

In this thesis, multiple approaches are used, each suited to the amount of flexibility expected, and indicated from simulations, for the given structure. For more information see the experimental section and the sections included in Chap. 3.

2.1.10 Error Analysis

Though widely regarded a disadvantage or limitation, the distance to intensity relationship of r^{-6} may actually be viewed as a clear advantage of the NOE over other 3D observables (*J*-coupling constants, RDCs etc.). This is due to the relationship between experimental error and resulting distance error, illustrated in Table 2.1.

This effectively means that rather large experimental errors, as a product of e.g. faulty assumptions of equal correlation times, spin diffusion or spectra with low *S*/*N*, will often result in quite small errors in the calculated distances, making the NOE approach highly favorable compared to other experimental methods. The difference on whether a too large or small NOE intensity is observed is due to the favoring of small distances from the r^{-6} dependency.

Table 2.1 Theoretical relationship between the error of the observed intensity and the error in the calculated distance both in %

Observed η error (%)	Resulting r error (%)	Observed η error (%)	Resulting r error (%)
−10	−1.8	10	1.6
−20	−3.8	20	3.0
−30	−6.1	30	4.3
−40	−8.9	40	5.5
−50	−12.2	50	6.5

Calculated using Eq. (2.3)

The decrease in NOE intensities for longer distances will result in longer distances being more prone to errors, since the NOE intensity will approach the spectral noise and the experimentally obtained data will more easily differ considerably from the actual NOE.

2.2 *J*-Coupling Constants

The next method used to obtain 3D structural information is by indirect dipole-dipole couplings, in the form of scalar or *J*-couplings. The phenomenon was first reported as "low beats" in an echo from spin echo methods for ethanol by Hahn in 1950 [28]. The origin of the observations was disputed until the Ramsey and Purcell formally introduced the *J*-coupling constant definition a couple of years later [29–31].

Scalar couplings are magnetic interactions transmitted by the bonding electrons by which the spins are indirectly connected [1, 32]. The coupling occurs as the magnetic moment of the nuclei polarizes electrons involved in the bond slightly, and this polarization is transmitted by overlapping orbitals to other nuclei. This leads to the spin state of one nucleus to influence the effective external magnetic field of neighboring nuclei and thus line splitting. The interaction is independent of the applied magnetic field since it only depends on the spin orientation and the orientation of the electrons which are paired due to the Pauli exclusion principle [6].

The situation is depicted in Fig. 2.7 over one bond for two bonded nuclei. The resulting signal will be a doublet due to the two different energy states, with a positive *J*-coupling constant [6]. The *J*-coupling constant is positive if the spin-spin coupling increases the energy when spin polarizations are parallel (triplet spin state) and decreases the energy when the spin polarizations are anti-parallel (singlet spin state) [6, 31]. Negative *J*-coupling constants have an opposite dependence. This assumes an equal sign of the gyromagnetic ratio of the two nuclei [6].

The energy of the spin-spin interaction, given as the Hamiltonian, depends on the magnetic moments and the position of nuclear spin vectors (I) of the involved nuclei as given in Eq. (2.22) [1, 6, 33].

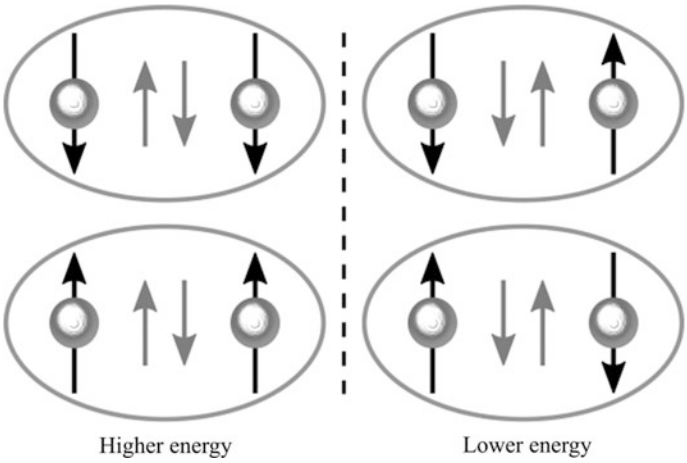


Fig. 2.7 Mechanism of scalar coupling between two neighboring spins (over one bond). The black arrows are the spin angular momentums and the grey arrows are the electron spins, paired due to the Pauli principle. The nuclear spins affect the spin of the electrons, and opposite polarizations of the nuclear spins are favored energetically as the electron spin polarizations are kept opposite. An equal nuclear spin polarization is high in energy as no favorable spin orientation may be achieved. Inspired by Levitt [6]

$$\hat{\mathcal{H}}_{IS}^{J} = J_{IS}\hat{\mathrm{I}}_I \cdot \hat{\mathrm{I}}_S \tag{2.22}$$

Implicit in the *J*-coupling constant is the product of the magnetogyric ratios of the involved nuclei and the reduced constant K is sometimes used, where this dependence is eliminated [1]. K is not used in this thesis.

Ramsey identified three interactions, which contribute to the *J*-coupling with the first being the most dominant: [31]

1. Fermi contact (FC) interaction of electronic and nuclear spins.
2. The electron orbital motion with nuclear spin.
3. Dipolar terms involving electronic and nuclear spins.

For the FC term which may be seen as a direct contact term of electrons and nuclei, σ electrons play a significant role in transmission while π electrons are mostly involved due to exchange interactions between the σ- and π-electronic systems [34].

The scalar couplings lead to the observed multiplets in e.g. 1D ^{1}H spectra and may be used in structure elucidation. Scalar couplings are also the foundation of most NMR experiments which employ coherent magnetization transfers e.g. COSY (mostly $^{2/3}J_{\mathrm{HH}}$), HSQC ($^{1}J_{\mathrm{CH}}$) or HMBC ($^{\mathrm{n}}J_{\mathrm{CH}}$) type experiments [1].

2.2.1 $^{3}J_{HH}$-Coupling Constants

The J_{HH}-coupling constants in the form of $^{n}J_{HH}$ (n = 2, 3), are arguably the easiest 3D structural parameter to extract as they are present in simple 1D ^{1}H spectra, and may be extracted when the resonances are not overlapping. If other types of coupling constants or a higher resolution, due to congested 1D spectra, are wanted, tailored experiments are usually needed; see Chap. 4 for a more elaborate discussion. The size of the J-coupling constants is, as mentioned above, based on the distance and degree of orbital overlap between the relevant nuclei. The influence of orbital overlap is the basis for the sinusoidal relationship between the size of coupling constants and a dihedral angle, which for $^{3}J_{HH}$-coupling constants are given by the semi-empirical Karplus or Karplus-like equations as illustrated in Fig. 2.8 [1, 35].

The constants present in the Karplus equations have generally been determined empirically from experimental data, though computational data may also be used, if e.g. less experimental data is present or to fully cover the given Karplus curve [35]. The relative size of the J-coupling constants is also dependent on the nuclei involved, e.g. HN–CH versus HC–CH coupling constants due to the difference in magnetic moment, and the properties of surrounding nuclei. This dependence is utilized in e.g. the Haasnoot, De Leeuw and Altona (HLA) or Altona equation [35–37]. Also the H–C–C internal angle, the C–C and N–H bond length and structural strain have been shown to influence the size of the coupling constants [6, 31, 38–40]. The Karplus equations generally take the form of Eq. (2.23) [35].

$$^{3}J_{HH}(\theta) = A\cos^{2}\theta + B\cos\theta + C \tag{2.23}$$

where θ is the dihedral angle and the constants A, B and C depend on the type of coupling as addressed above [35]. The equation was derived based on peptide moieties, and the amide version was used for NH–CH coupling constant calculations where A = 6.4, B = −1.4 and C = 1.9.

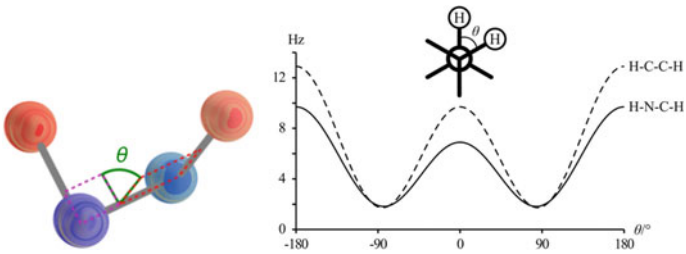
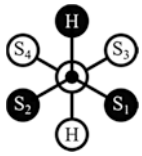


Fig. 2.8 Left: Dihedral angle between protons (red) three bonds apart, blue may be carbons or heteroatoms. Right: Dependence of ^{3}J-coupling constant on dihedral angle for peptide data as given by Karplus equations [35]

Table 2.2 Left: Constants for use in HLA calculations for different substitution patterns of the carbon atoms, Right: Relative orientation of substituents included in the HLA calculations

Type	P_1	P_2	P_3	P_4	P_5	P_6	P_7	
CH_3–CH_2R, CH_3–CHR_2, RCH_2–CH_2R	13.70	−0.73	0.00	0.56	−2.47	16.90	0.14	
R_2CH–CH_2R	13.22	−0.99	0.00	0.87	−2.46	19.90	0.00	
R_2CH–CHR_2	13.24	−0.91	0.00	0.53	−2.41	15.50	0.19	

An example of a more elaborate equation is the HLA equation given in Eq. (2.24), which was established to take the electronegativity of the neighboring nuclei into account [37].

$$
\begin{aligned}
{}^3J_{\mathrm{HH}}(\theta) &= P_1\cos^2\theta + P_2\cos\theta + P_3 + \sum \lambda_i\left(P_4 + P_5\cos^2(\varepsilon_i\theta + P_6|\lambda_i|)\right) \\
\lambda_i &= (X_\alpha - X_H) + P_7\sum\left(X_\beta - X_H\right)
\end{aligned}
\tag{2.24}
$$

This is the equation used for CH_n–CH_m coupling constants in this thesis unless stated otherwise, used with the appropriate constants from Table 2.2.

The post P_3 term is the sum of the attached groups and is dependent on the electronegativity (X) of the substituents (S_{1-4}), see Table 2.2 for clarifications. ε_i is equal to 1 for S_1 and S_3, and −1 for S_2 and S_4.

Unlike the more complicated averaging of NOEs presented in the previous section, *J*-coupling constants are usually averaged over multiple structures as the weighted average of the *J*-coupling constants for the different conformers. It is thus not averaged as the weighted average of the dihedral angle, an important distinction to make. The averaging of *J*-coupling constants may thus be considered orthogonal to the averaged NOEs.

2.3 Evaluating Data

To evaluate e.g. back-calculated distances from 3D structures to the relative distances obtained from NOEs, the mean absolute error (MAE) is used directly or on the percentage error, Eqs. (2.25) and (2.26). The MAE is also used for *J*-coupling constants.

$$
\mathrm{MAE} = \frac{1}{n}\sum\left|x_i^{exp} - x_i^{calc}\right| \tag{2.25}
$$

$$\mathrm{MAE}(\%) = \frac{1}{n}\sum \frac{\left|x_i^{exp} - x_i^{calc}\right|}{x_i^{exp}} \tag{2.26}$$

2.4 Introduction to Computational Chemistry

Often 3D structures are desired for discussing observation such as NMR data, biological assays or chemical properties in structural terms [41–43]. For the generation of optimized structures a plurality of methods are available [44]. In the literature multiple approaches have been proposed which may be used for structural calculation, and since the rationale behind the approaches taken in this thesis is discussed elsewhere in the text, only a short introduction is included here [44, 45]. Also, the theories, quantum chemistry and formula which govern the methods are only very briefly touched upon.

2.4.1 Molecular Mechanics and Dynamics

Molecular mechanics (MM) and molecular dynamics (MD) is a force field (FF) method, where the chemical bonds are described by classical mechanics where the positions of the nuclei are considered. MM is used in calculating the energy of a given system and e.g. differences in lengths (E_{str}), angles (E_{bend}) and dihedral angles (E_{tors}) between structures are scored in order to identify the lowest energy structure(s) [44]. Non-bonded interactions such as electrostatic (E_{el}) and van der Waals (E_{vdw}) interactions are also included. The energy of the structure will thus de- or increase by deviations in distances, angles etc., and the overall energy of the system will be determined by an energy calculation as the one given in (2.27) [44].

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \tag{2.27}$$

The term E_{cross} essentially covers combinations of the bonded interactions to further improve the force field when matched to experimental data. For example, the force field used in this thesis for organic solvents, MMFF, uses a combination of stretch and bend interactions for this term [44]. In the work of this thesis, the MM calculations were always coupled to conformational sampling, e.g. by Monte Carlo methods where the geometry is perturbed by varying e.g. torsion angles at random before minimization, or MD, see below [44, 45]. This results in optimized structures which generally should be good approximations to actual minima, but as each force field may score parameters differently, minima will be force-field dependent. It is thus not assumed that the relative energy between different compounds or conformers may be taken as accurate measures, and the structures are always compared to experimental data. It is thus important that the used methods result in

structures of the global as well as local minima, and that the conformational space is thoroughly sampled.

Also MD was used for simulations. In short, MD simulations treat the molecules as classical Newtonian systems, where the different interactions in between nuclei, such as bonds, produce a force on the nuclei. The system is then evolved for a given time, and the equations are continuously calculated in given time-steps [44, 45]. Since the time-steps are often short a great dependence on starting structure is often found for the simulations, and simulated annealing, where the temperature is varied (usually from high to low) may be needed, as the structures may traverse larger energy barriers [44].

The solvents are included in the simulations, since the experimental properties observed by NMR in this thesis are observed in the liquid state. The treatment of solvents in MM calculations, as well as Density Functional Theory (DFT) calculations is complex, and will not be discussed. Briefly, the solvent implementation may be either implicit, where the solvent is treated as a continuous medium, usually where the properties of the solvent is given by a dielectric constant that interacts with the molecule, or explicit where multiple solvent molecules are included in MD simulations and evolved with the molecule of interest [44, 45].

2.4.2 Density Functional Theory

Density Functional Theory (DFT) is a computational approach for calculation of electron energies. Here, approximations are needed as the electron energy cannot be determined or estimated by classical mechanics, and a set of functions is used to determine the energy of a system by the electron density of said system [44, 45].

The basis of DFT calculations is formed by the functionals which are the series of formulae used to describe and solve the electron density. Multiple functionals have been derived and are available, often from a combination of different exchange and correlation functionals [44]. The DFT functionals used in this study are the Becke Three Parameter Hybrid Functionals (B3LYP), probably the most widely used method for small molecules, and MPW1PW91, which is a single parameter functional developed in part for NMR calculations [44, 46]. Both are hybrid functionals and include a mixture of Hartree-Fock exchange and DFT exchange-correlation [44].

The functionals use basis sets which may also be varied according to the investigated problem. A basis set consists of predetermined functions which are used to describe molecular orbitals centered at each nuclei using a linear combination of Gaussian type orbitals (GTOs) [44]. The GTOs are used to approximate Slater-type orbitals (STOs) by linear combinations. STOs are not determined directly due to computational inefficiency. The basis set may be augmented by polarization functions or diffuse functions. Polarization functions add another orbital to the initially used orbitals for a given nucleus. For a nucleus with and outer *p*-orbital a *d*-orbital may thus be added, which allows for more asymmetry around the nucleus.

Diffuse functions are Gaussian functions added to better describe the electron density far from the nucleus [44].

When applying DFT calculations to structural challenges the Pople basis set was used throughout this thesis. These basis sets are built as exemplified for 6–31 g(d); the core orbitals are described by six GTOs, the inner part of the valence orbitals are described by three GTOs and the outer part of the valence orbitals by one GTO. The (d) means that a polarization function is used and a *d*-type polarization function is added to the heavy atoms. Other types of basis sets could be the correlation consistent basis sets e.g. cc-pVDZ. The correlation consistent basis sets differ from the Pople set as they are designed to converge towards the basis set limit, where properties are described as if using an infinite basis. As such a greater basis set should always lead to better calculations of the property of interest. It was recently reported that the basis set limit, where the calculated properties converge, was reached for the calculation of NMR properties, which should lead to an even stronger coupling of experimental NMR spectroscopy and computational chemistry [47].

It is important to choose a basis set that serves the need of the system in question. Larger basis sets may lead to increase in accuracy if the orbitals are better described but will also increase the computational time. The increase in accuracy is not always realized though as an observable may be better parameterized by a smaller basis set when using the Pople basis sets. It is thus often useful to screen different DFT functionals and basis sets in order to obtain a match for a given challenge, or to take inspiration for methods which have been shown to work well for small molecules by others [42, 48, 49]. For the optimization of structures it is always useful to slowly increase the basis set as the computational time increases with size starting from e.g. output from FF methods. The computation time may thus be minimized by submitting already optimized structures to the more expensive calculations. After optimization the resulting structures were used for the calculations of NMR observable properties, namely chemical shifts and *J*-coupling constants. Whether it is needed to optimize structures to a DFT-level of theory for chemical shift calculations has been questioned, but it seems to be needed for calculations of *J*-coupling constants [49–53].

2.4.3 Chemical Shifts

The chemical shift is defined as the difference in shielding tensors of nuclei compared to the shielding tensor of a reference, which for ^{1}H and ^{13}C nuclei are the shielding tensors of tetramethylsilane (TMS). Similarly the calculated property from DFT calculations is the NMR shielding tensor, which is a second order magnetic property, defined as the energy derivate of the nuclear magnetic moment μ_I with respect to the applied magnetic field, B [47].

$$\sigma_I = \partial^2 E / \partial \mu_I \partial B \big|_{B,M_I=0} \tag{2.28}$$

The calculations of shielding tensors are based on the interaction of the applied magnetic field, the magnetic field generated by the nuclei and the magnetic field generated by the rapid movement of the electrons in a molecule [42, 49, 53, 54].

The first two terms are termed static fields and perturb the kinetic energy of the Hamiltonian so that if a finite basis set is used to model the electron distribution, the operator depends on the origin of the coordinate system known as the gauge. This is circumvented by either the GIAO (gauge invariant atomic orbitals) method, which uses basis functions that have an explicit field dependence, or the CSGT (continuous set of gauge transformation) which performs a gauge transformation at each point in space. It has been shown that for the nuclei involved in NMR the two methods work equally well, but that the GAIO method converges faster [42].

Generally two methods are available for correlating experimental chemical shifts to the calculated GIAO nuclear magnetic shielding tensors. One is to calculate the tensors for TMS and use these tensors as reference in comparison to an internal standard for liquid NMR. There is a possible caveat as the GIAO tensors of TMS must be calculated for each functional and basis set used, and the silicon in TMS is not necessarily parameterized well in all of the used basis sets, a problem especially for smaller basis sets [42]. The second, and more general, approach is to linearly correlate the shielding tensors to the observed chemical shifts, as in Eq. 2.29 [49, 53].

$$\delta_{scaled} = \frac{\sigma_{GIAO} - b}{a} \tag{2.29}$$

where a is the slope and b the intercept of the correlation of observed shifts and calculated shielding tensor. The approach is simpler and less computationally demanding but could potentially result in overfitting of the data, as each structure is individually scaled to experimental data. The approach is used in this thesis in line with most literature [42, 49, 53].

2.4.4 J-Coupling Constants

Nuclear spin-spin coupling constants between two nuclei are calculated by the second derivatives of the energy with regard to their magnetic moments as seen in Eq. (2.30) [34, 48].

$$J_{IS,xy} = \partial^2 E / \partial \mu_{I,x} \partial \mu_{S,y} \tag{2.30}$$

As mentioned in an earlier section the coupling of the nuclei is comprised of the sum of contributions of different terms covering different mechanisms: The spin hyperfine interaction and the spin-orbit (SO) coupling [34, 48]. The SO interactions

may be divided into a diamagnetic (DSO) and a paramagnetic (PSO) component and depend on the magnetic moments created by the movement of electrons. The hyperfine interactions are comprised of the Fermi contact (FC) term and the spin-dipole (SD) operator and depend on the interaction of an electron's spin with the local magnetic field of the nuclei [48]. The FC term describes the interaction at the nucleus and is usually dependent on the nature and environment of the nuclei, by far the largest contribution to the total coupling constant. For proton-proton coupling constants all other terms than the FC may be neglected, and the result is scaled to correct for this and other possible inherent errors, an approach which has been shown to work well for small molecules [48]. The coupling constants are returned in Hz and are thus immediately comparable to the experimental data. Unlike shielding, scalar coupling calculations will be very sensitive to the shape of the electronic wave-functions near the nucleus, which should be modelled as accurately as possible for the FC term. The calculations are thus more demanding than the chemical shift calculations, as this is hard to simulate using GTOs [34]. The methods used for the calculations of *J*-coupling constants are discussed in the text in Chap. 4.

References

1. H. Günther, *NMR Spectroscopy*, 3rd edn. (Wiley-VCH, Weinheim, 2013)
2. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 1st edn. (VCH Publishers, Weinheim, 1989)
3. A.W. Overhauser, Phys. Rev. **89**, 689–700 (1953)
4. A.W. Overhauser, Phys. Rev. **92**, 411–415 (1953)
5. I. Solomon, Phys. Rev. **99**, 559–565 (1955)
6. M.H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd edn. (John Wiley & Sons Ltd, Chichester, 2008)
7. A.D. Bax, D.G. Davis, **63**, 207–213 (1985)
8. A. Bax, S. Grzesiek, in *Encyclopedia of Magnetic Resonance, Online*, (John Wiley & Sons, Ltd, 2007), pp. 1–10
9. T.-L. Hwang, A.J. Shaka, J. Am. Chem. Soc. **114**, 3157–3159 (1992)
10. A. Kjellberg, G. Widmalm, Biopolymers **50**, 391–399 (1999)
11. N. Bross-Walch, T. Kühn, D. Moskau, O. Zerbe, Chem. Biodivers. **2**, 147–177 (2005)
12. S. Macura, B.T. Farmer II, L.R. Brown, J. Magn. Reson. **70**, 493–499 (1986)
13. N.H. Andersen, H.L. Eaton, X. Lai, Magn. Reson. Chem. **27**, 515–528 (1989)
14. T.L. James, Curr. Opin. Struct. Biol. **1**, 1042–1053 (1991)
15. C.E. Keller, W.R. Carper, Magn. Reson. Chem. **31**, 566–572 (1993)
16. C.P. Butts, C.R. Jones, E.C. Towers, J.L. Flynn, L. Appleby, N.J. Barron, Org. Biomol. Chem. **9**, 177–184 (2011)
17. B.A. Borgias, M. Gochin, D.J. Kerwood, T.L. James, Prog. Nucl. Magn. Reson. Spectrosc. **22**, 83–100 (1990)
18. H. Hu, K. Krishnamurthy, J. Magn. Reson. **182**, 173–177 (2006)
19. C.N. Chi, D. Strotz, R. Riek, B. Vögeli, J. Biomol. NMR **62**, 63–69 (2015)
20. B.A. Borgias, T.L. James, Methods Enzymol. **176**, 169–183 (1989)
21. B.A. Borgias, T.L. James, J. Magn. Reson. **87**, 475–487 (1990)
22. C.B. Post, J. Mol. Biol. **224**, 1087–1101 (1992)

23. C.R. Jones, C.P. Butts, J.N. Harvey, Beilstein J. Org. Chem. **7**, 145–150 (2011)
24. C.P. Butts, C.R. Jones, Z. Song, T.J. Simpson, Chem. Commun. **48**, 9023–9025 (2012)
25. C.P. Butts, C.R. Jones, J.N. Harvey, Chem. Commun. **47**, 1193–1195 (2011)
26. R. Bürgi, J. Pitera, W.F. van Gunsteren, J. Biomol. NMR **19**, 305–320 (2001)
27. X. Daura, I. Antes, W.F. van Gunsteren, W. Thiel, E. Mark, Proteins Struct. Funct. Genet. **36**, 542–555 (1999)
28. E.L. Hahn, Phys. Rev. **80**, 580–594 (1950)
29. N.F. Ramsey, E.M. Purcell, Phys. Rev. **85**, 143–144 (1952)
30. N.F. Ramsey, Phys. Rev. **91**, 303–307 (1953)
31. M. Barfield, in *Encyclopedia of Nuclear Magnetic Resonance*, ed. by D.M. Grant, R.K. Harris (John Wiley & Sons, Ltd, Chichester, 1996), pp 2520–2530
32. J. Keeler, in *Understanding NMR Spectroscopy*, 2nd edn. (John Wiley & Sons, Ltd, 2010)
33. T. Engel, *Quantum Chemistry & Spectroscopy*, 1st edn. (Pearson Education Inc, San Francisco, 2006)
34. I. Alkorta, J. Elguero, Int. J. Mol. Sci. **4**, 64–92 (2003)
35. S. Karplus, M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **69**, 3204–3206 (1972)
36. C. Altona, C.A.G. Haasnoot, Org. Magn. Reson. **13**, 417–429 (1980)
37. C.A.G. Haasnoot, F.A.A.M. DeLeeuw, C. Altona, Tetrahedron **36**, 2783–2792 (1979)
38. P.A. Karplus, Protein Sci. **5**, 1406–1420 (1996)
39. F. Li, J.H. Lee, A. Grishaev, J. Ying, A. Bax, Chem. Phys. Chem. **16**, 572–578 (2015)
40. B. Coxon, Adv. Carbohydr. Chem. Biochem. **62**, 17–82 (2009)
41. G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, Chem. Rev. **107**, 3744–3779 (2007)
42. M.W. Lodewyk, M.R. Siebert, D.J. Tantillo, Chem. Rev. **112**, 1839–1862 (2012)
43. J. Klages, M. Coles, H. Kessler, Anal. **132**, 692–705 (2007)
44. F. Jensen, in *Introduction to Computational Chemistry*, 1st edn. (John Wiley & Sons, Ltd, 1999)
45. A.R. Leach, *Molecular Modelling—Principles and applications*, 2nd edn. (Essex, England, Pearson Education Limited, 2001)
46. DFT Gaussian, http://www.gaussian.com/g_tech/g_ur/k_dft.htm. Accessed 14 Nov 2015
47. S.R. Jensen, T. Flå, D. Jonsson, S. Monstad, Phys. Chem. Chem. Phys. (2016)
48. T. Bally, P.R. Rablen, J. Org. Chem. **76**, 4818–4830 (2011)
49. R. Jain, T. Bally, P.R. Rablen, J. Org. Chem. **74**, 4017–4023 (2009)
50. A. Bagno, F. Rastrelli, G. Saielli, Chem. Eur. J. **12**, 5514–5525 (2006)
51. M.G. Chini, R. Riccio, G. Bifulco, Magn. Reson. Chem. **46**, 962–968 (2008)
52. S.G. Smith, J.M. Goodman, J. Org. Chem. **74**, 4597–4607 (2009)
53. S.G. Smith, J.M. Goodman, J. Am. Chem. Soc. **132**, 12946–12959 (2010)
54. D.A. Forsyth, A.B. Sebag, J. Am. Chem. Soc. **119**, 9483–9494 (1997)

Chapter 3
Application of NOEs and $^3J_{HH}$-Couplings in 3D Structure Determination

3.1 Natural Products

The structures of homomorphosin A-F were elucidated during work done on the Master's Thesis "Chemistry of Black Aspergilli" by the author. A description of the purification and elucidation of the compounds is thus not included here. The elucidation of the 3D structural features of homomorphosin A and homomorphosterol was achieved during the timeframe of this thesis. The structural elucidation and 3D structural calculations of the cyclomorhosins were done entirely during the timeframe of this thesis.

3.1.1 Introduction to Natural Products

Natural products are the collected term of metabolites from natural sources, usually microbes (fungi or bacteria) or plants. Metabolites may be primary, those crucial for the growth of the organism and thus shared by many species, or secondary metabolites (SMs), where the latter are generally small to medium sized molecules with another purpose than growth [1, 2]. The functions of the SMs are not always known, but e.g. defence mechanisms, attractants or colouring agents have been suggested [3, 4]. The discovery of penicillin in 1929 spurred the interest of in SMs due to their applicability as drugs and still the majority of approved drugs are of biological origin or derivatives of natural products [5, 6].

The SMs are divided into classes as polyketides, terpenes, non-ribosomal peptides and alkaloids, based on the origin of the precursors for the SMs and the genes utilized in the biosynthesis [1]. The biosynthesis is catalysed by enzymes such as polyketide synthases and non-ribosomal peptide synthetases, where covalent acyl-enzyme intermediates are formed and followed by condensations and tailoring to give the desired product, which are often enantiomerically pure compounds with

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_3

many stereocenters and complex ring systems [2, 7, 8]. The molecules, and derivatives thereof, have usually been synthesised by total synthesis for further investigations, but the emerge of synthetic biology, where the machinery of the microbes are used to biosynthesise the wanted compound is a possible future in drug discovery [9].

3.1.2 *Aspergillus Homomorphus*

Aspergillus homomorphus (*A. homomorphus*) is a fungus residing in the *Aspergillus* section *Nigri* (black aspergilli), seen in Fig. 3.1 [10, 11]. It is geographically highly specific and has only been identified from soil in the Dead Sea area and from an Argentinian vineyard [12, 13]. It is thus a relatively rare fungus and the metabolic profile was previously only sparsely described [14].

3.1.3 *Homomorphosin A-F*

The homomorphosins A-F are depicted in Fig. 3.2. They are all diketopiperazines made from tryptophan and either valine, 2-aminobutyrate or alanine, and attached isoprene units.

Homomorphosin A is the most predominant metabolite produced by *A. homomorphus* on any growth medium, and was subjected to a 3D structural analysis to verify the structure and to obtain structural insights. The calculations may also serve as an introduction to the techniques utilized throughout this chapter, from the simplicity of the structure. Distances from NOEs were obtained as described in the theory, by relative intensities in the linear range. The solvent was DMSO-d_6.[1]

Using a single structure approach, the best fit of the experimental data and a 3D structure is seen in Table 3.1 for NOEs and in Table 3.2 for *J*-coupling constants. Major violations in the NOE derived distances were observed for the flexible parts of the molecule which needed to be averaged.

If multiple structures were used the correlation between the NMR and structural data was much better, and it was verified that the observed differences in distances for a single structure originated from structural flexibility, see Tables 3.3, 3.4 and Fig. 3.3. The utilization of multiple structures allowed the assignment of the "pro-chiral" methyl groups 18 and 19 as well as the diastereotopic protons 25 from an increase in MAE upon inversion, see Appendix A4. A constraint in the rotational space of the valine side-chain was initially suggested due to the low H_α-H_β coupling constant and the difference in the methyl chemical shifts, which was

[1]Note that the solvents when acquiring NMR experiments throughout this thesis are always deuterated, unless explicitly stated otherwise.

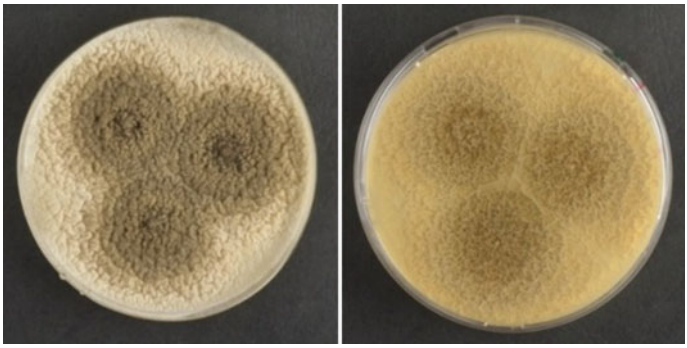


Fig. 3.1 *A. homomorphus* fungi grown for 7 days on YES media, left is top- and right bottom view

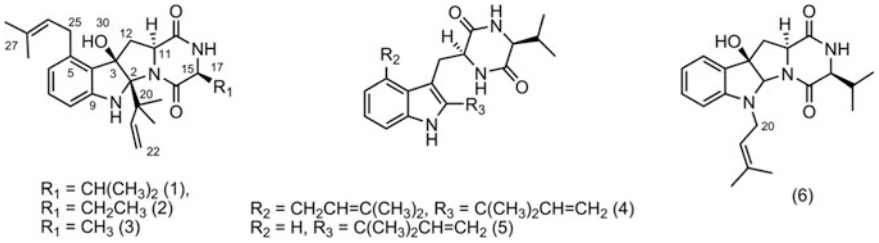


Fig. 3.2 Novel NRPs from *A. homomorphus*; Homomorphosin A-F (1–6). Homomorphosin A (1) is the main secondary metabolite produced by *A. homomorphus*

confirmed from the NOE distances. The methyl groups 23/24 were not solved due to similar and slightly overlapping intensities, and they are treated as the summed intensities in the calculations.

In conclusion the 3D structural investigation confirmed the structure obtained from other NMR experiments including the usage of qualitative NOEs. This is not surprising due to the rigid structure, but still the ability to determine the distances reliably dismisses any doubts. Also the accuracy is a good example of the degree of correlation between experimental and theoretical data that may be envisioned for the following small molecules, including the fact that rotatable bonds will be harder to fit than rigid parts, which is not surprising. Still, a pretty good fit of the experimental data was obtained from groups that exhibit freely rotatable bonds, and it is shown that allowing multiple conformers leads to a better fit of these parts, enabling supported suggestions to be made regarding stereochemistry. The ability to average over multiple conformers will be needed in the following sections.

The absolute stereochemistry was solved by Marfey's reaction as the valine was L-valine. Thus all stereocenters could be solved in relation.

Table 3.1 Comparison of the distances from NOE intensities and 3D structures using a single structure, with the lowest MAE, optimized to B3LYP/6-31 g(d) (τ_m = 200 ms)

H1	H2	Exp. dist.	Lower bound	Violation	Upper bound	Violation
1	**8**	**2.66**	**2.39**		**2.92**	
1	21	2.88	4.18	1.30	5.11	
1	23/24	2.44	2.10		2.56	
6	25a	2.78	2.70		3.30	
6	26	2.92	2.71		3.31	
6	29	3.67	3.32		4.05	
7	**8**	**2.48**[a]	**2.23**		**2.73**	
11	**15**	**2.46**	**2.20**		**2.69**	
14	**15**	**2.72**	**2.46**		**3.00**	
14	17	2.94	3.33	0.39	4.07	
15	18	3.05	3.17	0.12	3.88	
15	19	2.58	2.25		2.75	
15	17	2.53	2.27		2.77	
17	**18**	**2.30**	**2.08**		**2.54**	
17	**19**	**2.30**	**2.07**		**2.53**	
21	23/24	2.47	2.26		2.77	
21	12a	2.87	2.44		2.98	
21	30	2.80	2.96	0.16	3.62	
22b	23/24	2.28	2.07		2.53	
22b	18	4.30	3.69		4.51	
25b	29	2.59	2.46		3.00	
25b	12b	2.26	1.88		2.30	
26	**28**	**2.39**	**2.15**		**2.63**	
30	23/24	2.41	2.38		2.91	
30	12a	2.84	1.92		2.35	0.49
30	25b	2.95	2.89		3.53	

All distances in Å. Lower and upper bounds are defined as ±10% of the distances from the 3D structures. Rigid distances are in bold. [a]Reference distance used

Table 3.2 Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation from 3D structures using best fit single structure, in Hz

H1	H2	*J* exp.	*J* calc.	Diff
11	12b	7.3	8.0	0.7
15	17	2.0	2.3	0.3
11	12a	11.5	10.3	1.2

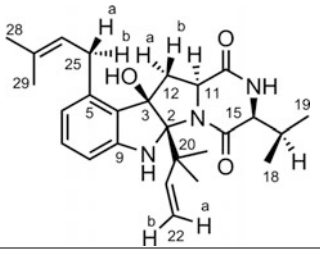


Table 3.3 Comparison of the distances from NOE intensities and 3D structures using multiple structures from MM (τ_m = 200 ms)

H1	H2	Exp. dist.	3D Dist.	Diff (Å)	Diff (%)
1	**8**	**2.61**	**2.66**	**0.05**	**1.9**
1	21	2.88	2.96	0.08	2.7
1	23/24	3.34	3.37	0.03	0.9
6	25a	2.78	2.76	0.03	0.9
6	26	2.79	2.80	0.01	0.3
6	29	4.41	4.42	0.01	0.1
7	**8**	**2.48**[a]	**2.48**	**0.00**	**0.0**
11	**15**	**2.53**	**2.44**	**0.09**	**3.7**
14	**15**	**2.72**	**2.73**	**0.00**	**0.1**
14	17	2.94	2.93	0.01	0.5
15	18	3.42	3.50	0.08	2.3
15	17	2.25	2.52	0.27	12.1
15	19	2.76	3.01	0.25	9.1
17	**18**	**2.81**	**2.62**	**0.20**	**7.0**
17	**19**	**2.75**	**2.61**	**0.14**	**5.0**
21	12a	2.87	2.96	0.09	3.1
21	23/24	3.10	3.28	0.18	5.8
21	30	2.77	2.89	0.12	4.2
22b	18	4.71	4.67	0.04	0.8
22b	23/24	3.05	3.16	0.11	3.6
25b	12b	2.18	2.27	0.09	4.1
25b	29	3.11	3.07	0.04	1.3
26	**28**	**2.96**	**2.82**	**0.14**	**4.6**

(continued)

Table 3.3 (continued)

H1	H2	Exp. dist.	3D Dist.	Diff (Å)	Diff (%)
30	12a	2.84	2.75	0.09	3.1
30	23/24	3.09	3.35	0.26	8.5
30	25b	2.81	2.92	0.11	4.0
			MAE	0.10	3.5

All distances in Å. [a]Reference

Table 3.4 Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation from 3D structures using multiple structures from MM, in Hz

H1	H2	*J* exp.	*J* calc.	Diff
11	12b	7.3	6.5	0.8
11	12a	11.5	10.2	1.3
15	17	2.0	2.0	0.0

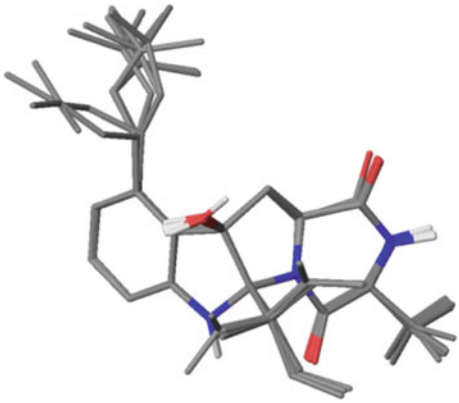


Fig. 3.3 Representation of the structural space inhabited by homomorphosin A, suggested from NOE and *J*-coupling constant data. The heavy atoms in the tetracyclic system of the 10 most highly populated conformers are overlain. Only the hydrogens of polar groups are shown

3.1.4 Cyclomorphosins

Two compounds were isolated with the molecular masses $[M + H]^+$ of 850.3307 and 866.3611 m/z, determined from HRMS analyses. This translated into multiple possible molecular formulae for each structure, as the number of possibilities increase with size due to an increase in possible nuclei and associated isotope patterns. The most probable constituent formulae were $C_{46}H_{43}N_9O_8$ or $C_{45}H_{47}N_5O_{12}$ and $C_{47}H_{47}N_9O_8$ or $C_{46}H_{51}N_5O_{12}$, respectively for the two masses. In each case the latter constituent formulae were identified as the correct ones from NMR spectroscopy.

Structural elucidation

The structural elucidation process is exemplified for cyclomorphosin A (CM-A), and a similar approach was used to elucidate the structure of cyclomorphosin B (CM-B), but from a significantly lower amount of compound. The structures of both NRPs are shown in Fig. 3.4.

The 1D ^{1}H spectrum of CM-A is found in Fig. 3.5. The difficulty of elucidation was increased by the presence of a major and a minor conformer in the ratio 3:1 as illustrated in the inset of Fig. 3.5. For CM-B the ratio was determined to be

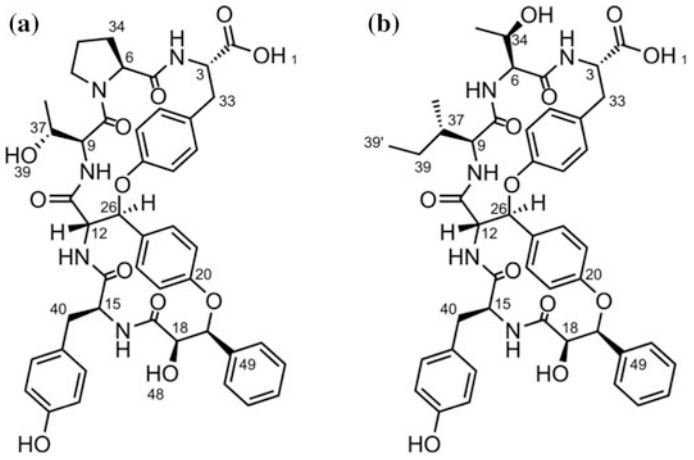


Fig. 3.4 Novel NRPs from *A. homomorphus*: Cyclomorphosin A (left) and B (right)

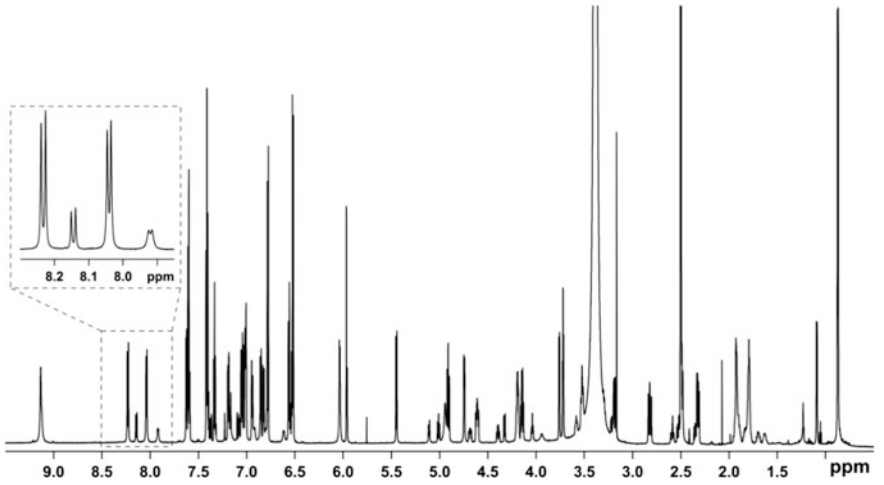


Fig. 3.5 1D spectrum of cyclomorphosin A (CM-A) in DMSO-d_6 at 800 MHz. A carboxylic acid resonance at 12.76 ppm is omitted. Residual solvent and water peaks are cut off. Inset: Two slowly exchanging conformers of CM-A

approximately 10:1. The appearance of the resonances is the result of a second conformational species being present which interchanges slowly with the primary conformational species on the NMR chemical shift time scale. Luckily, only few resonances overlap between the conformers making it possible to unambiguously assign the shifts to each conformer.

The amino acid residues were identified as distinct spin systems from DQF-COSY and HSQC-TOCSY spectra, and these spin systems were connected by HMBC correlations and qualitative NOEs as indicated in Fig. 3.6. A very low *J*-coupling constant between H-18 and H-19 made this connection difficult to assign due to a lack of correlations in e.g. the DQF-COSY experiment. The joint was thus determined from HMBC correlations, NOEs and a process of elimination of possible connectivities of the nuclei. The data is summarized in Table 3.5 and Appendix A4.

The two NRPs are each made of five amino acids and a building block which may originate from an intermediate in the shikimate pathway [15]. Both structures exhibit a bi-cyclic motif, Fig. 3.7, none of which have previously been reported from any fungal sources. Similar structures to the 14-membered ring (A), from different sets of amino acids, have previously been reported from plants [16–20]. The 17-membered ring (B) is a completely novel motif from natural sources. Both ring structures were rather rigid due to the rigidity of aromatic systems and amide bonds, especially the 14-membered ring with a smaller ring size.

The structures of the cyclomorphosins have nine chiral centers in the structure. Multiple examples of D-amino acids have been reported in NRPs, and the stereochemistry may not simply be ascribed to L-amino acids [8, 21]. The relative orientation of the chiral centers was solved by first decimating the amount of possible diastereomers from qualitative NOE analysis, applying distances obtained from relative NOE intensities to the remaining diastereomers and identifying the best fit to the data.

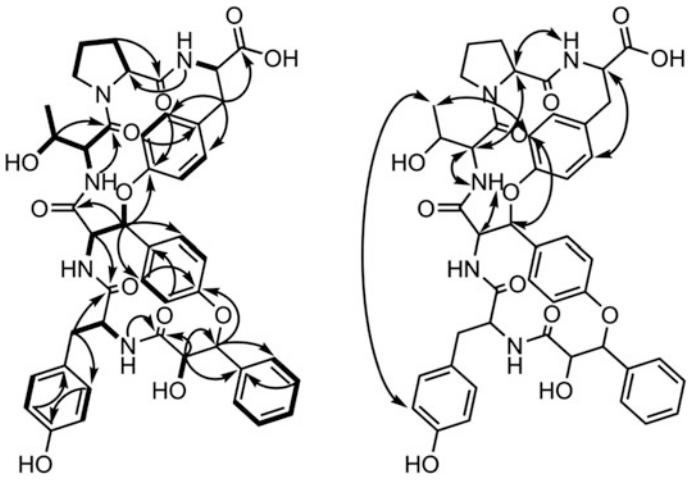


Fig. 3.6 Left: DFQ-COSY and HSQC-TOCSY (▬) spin systems and key HMBCs (→) of CM-A. Right: Key NOEs of the B ring of CM-A

Table 3.5 Chemical shifts and HMBC correlations, assigned for CM-A and CM-B in DMSO-d_6

	CM-A			CM-B		
#	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC
1	12.76 (1H, s)	–		ND	–	
2	–	172.6		–	ND	
3	4.60 (1H, ddd, 11.5, 8.5, 4.7)	51.0	2, 33	4.50 (1H, td, 10.1, 3.6)	52.2	
4	8.03 (1H, d, 8.7)	–	5, 33	7.83 (1H, d, 9.4)	–	
5	–	170.6		–	168.3	
6	4.75 (1H, d, 6.7)	58.9	34, 35, 36	4.19 (1H, dd, 8.0, 6.2)	57.5	5, 8, 34, 35
7	–	–		6.87 (1H, d, 6.8)	–	
8	–	165.8		–	170.0	
9	3.73 (1H, t, 4.4)	55.9	8, 37, 38	4.13 (1H, t, 8.7)	58.1	8, 37, 39
10	7.60 (1H, d, 4.6)	–	8, 9, 11	8.53 (1H, d, 9.4)	–	11
11	–	168.2		–	169.2	
12	4.90 (1H, t, 9.6)	58.0	11, 14, 26	4.66 (1H, dd, 9.7, 7.3)	58.6	14, 26
13	8.21 (1H, d, 10.2)	–	14	8.33 (1H, d, 9.9)	–	14
14	–	169.4		–	169.3	
15	4.14 (1H, dt, 9.2, 7.9)	52.5	14, 40	4.39 (1H, dt, 9.3, 8.0)	51.7	17
16	6.56 (1H, d, 9.1)	–	15, 17	6.48 (1H, d, 9.5)	–	
17	–	169.9		–	169.1	
18	3.75 (1H, d, 5.5)	74.1	17, 49	3.75 (1H, mult)	74.2	
19	5.95 (1H, s)	78.2	17, 18, 20, 49, 52	5.93 (1H, s)	78.7	49, 50/52
20	–	155.7		–	155.6	
21	7.04 (1H, mult)	116.3	23, 25	7.04 (1H, mult)	116.9	20, 25
22	7.19 (1H, dd, 8.1, 2.6)	129.6	20, 24, 26	7.01 (1H, mult)	128.9	24
23	7.02 (1H, mult)	120.3	21, 25	7.06 (1H, d, 8.5)	120.6	21, 25

(continued)

Table 3.5 (continued)

	CM-A			CM-B		
#	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC
24	7.42 (1H, mult)	127.6	20, 22, 26	7.46 (1H, d, 8.5)	126.8	22
25	–	131.6		–	132.2	
26	5.44 (1H, d, 8.9)	79.4	11, 12, 22, 24, 25, 27	5.53 (1H, d, 7.2)	78.9	11, 22, 24, 25, 27
27	–	155.0		–	155.2	
28	6.82 (1H, dd, 8.2, 2.4)	118.8	27, 30, 32	6.76 (1H, d, 6.8)	ND	
29	7.01a (1H, mult)	127.8	27, 31	6.98 (1H, mult)	ND	
30	7.01b (1H, mult)	113.8	28, 32	6.76 (1H, d, 6.8)	ND	
31	6.93 (1H, dd, 8.3, 1.1)	129.5	27, 29, 33	6.98 (1H, mult)	ND	
32	–	129.6		–	ND	
33a	2.82 (1H, dd, 14.3, 11.9)	34.9	3, 29, 32	2.54 (1H, mult)	36.4	
33b	3.18 (1H, dd, 14.3, 4.1)	34.9	3, 29, 32	3.09 (1H, dd, 13.3, 3.6)	36.4	
34a	1.92 (1H, mult)	20.8	5, 35, 36	3.75 (1H, mult)	65.8	8
34b	1.79 (1H, mult)	20.8	5, 6, 341	–	–	
35a	1.92 (1H, mult)	30.3	5, 34, 36	0.95 (3H, d, 6.2)	19.7	6, 34
35b	1.79 (1H, mult)	30.3	5, 6, 362	–	–	
36a	3.38 (1H, mult)	45.9	34	4.61 (1H, d, 5.1)	–	34, 35
36b	3.52 (1H, mult)	45.9	34, 35	–	–	
37	4.19 (1H, mult)	64.8	8	1.66 (1H, mult)	36.5	
38	0.85 (3H, d, 6.4)	17.4	9, 37	0.75 (1H, d, 6.7)	15.1	9, 34, 39
39a	4.93 (1H, s)	–		0.98 (1H, mult)	24.4	
39b	–	–		1.30 (1H, mult)	24.4	
39’	–	–		0.74 (1H, d, 7.3)	15.1	37, 39
40a	2.31 (1H, td, 13.7, 7.0)	38.3	14, 15, 41, 42	2.33 (1H, d, 13.9, 7.2)	37.9	14, 15, 41, 42/44

(continued)

Table 3.5 (continued)

	CM-A			CM-B		
#	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC	δ_H (ppm) (Int, mult, *J* [Hz])	δ_C (ppm)	HMBC
40b	2.49 (1H, mult)	38.3	14, 15, 42	2.49 (1H, mult)	37.9	41, 42/44
41	–	126.5		–	126.5	
42	6.78 (2H, d, 8.4)	129.5	40, 41, 42/44, 43/45, 46	6.80 (1H, d, 8.4)	129.6	40, 42/44, 46
43	6.51 (1H, mult)	114.4	41, 43/45, 46	6.51 (1H, d, 8.3)	114.4	41, 43/45, 46
44	6.78 (2H, d, 8.4)	129.5	40, 41, 42/44, 43/45, 46	6.80 (1H, d, 8.4)	129.6	40, 42/44, 46
45	6.50 (1H, mult)	114.4	41, 43/45, 46	6.51 (1H, d, 8.3)	114.4	41, 43/45, 46
46	–	155.5		–	155.4	
47	9.13 (1H, s)	–	43/45	9.08 (1H, s)	–	43/45, 46
48	6.02 (1H, d, 4.5)	–		6.01 (1H, d, 6.6)	–	17, 18, 19
49	–	140.4	17, 19	–	140.6	
50	7.59 (2H, mult)	126.5	19, 54	7.59 (1H, d, 7.6)	126.6	19, 50/52, 54
51	7.41 (1H, mult)	127.7	49	7.40 (1H, t, 7.5)	127.9	49, 51/53
52	7.59 (2H, mult)	126.5	19, 54	7.59 (1H, d, 7.6)	126.6	19, 50/52, 54
53	7.39 (1H, mult)	127.7	49	7.40 (1H, t, 7.5)	127.9	49, 51/53
54	7.33 (1H, t, 7.2)	127.3	50/52	7.32 (1H, t, 7.3)	127.2	50/52

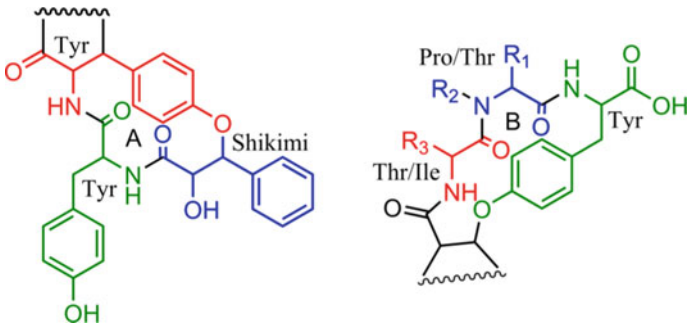


Fig. 3.7 Motif of the 14- (A) and 17-membered ring (B) of cyclomorphosin A and B. Colors illustrate the different putative building blocks

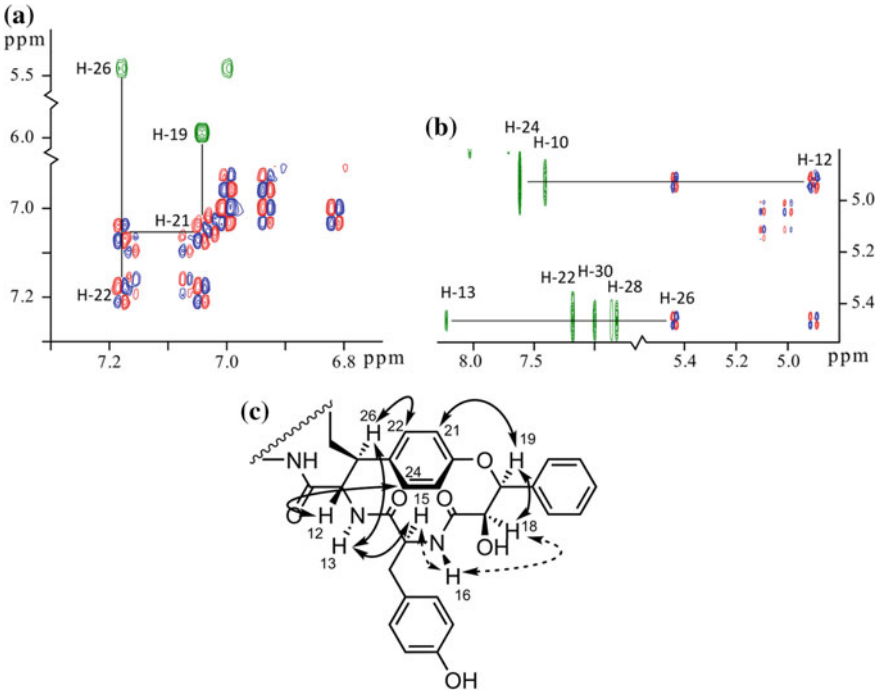


Fig. 3.8 Above: DQF-COSY (red and blue) and NOESY (green) spectra of CM-A. The spectra illustrates that H-26 and H-19 are situated on the same face (**a**) of the structure, while H-26 and H-12 are situated on opposite faces (**b**). Note that parts of the spectra, along the indirect axis in the left spectrum and the direct axis in the right, have been removed for convenience. **c** Key NOEs of the A ring of CM-A. The orientation indicated for amide protons shows the orientation of the protons compared to the face of the ring. Dashed arrows are weak NOEs

Aspects of the relative stereochemistry in ring A around the oxidized tyrosine was directly read from qualitative NOEs as illustrated in Fig. 3.8. NOEs indicate that H-26 and H-19 are positioned on the same face of the structure, while H-26 and H-12 are positioned on opposite faces, resulting in very different NOEs for these neighboring protons.

The differences in observed NOEs between H-12 and H-26 may not be proof of the relative stereochemistry, but considering a rather large *J*-coupling constant of 9.6 Hz between the two protons, a conformation displaying the observed NOEs while the protons are on the same face of the rings is impossible. From the HLA equation, the dihedral angles that result in a coupling constant of 9.6 Hz are ±3° and ±156°. The first is not compatible with the NOE data, the latter is only possible if on opposite faces of the rings. It was also realized through computational simulations that the angle was rarely near 0° for the threo structures while it was often near ±155° for the erythro conformations.

In a similar fashion, it was realized that the H-19 and H-26 had to be on the same face of the ring. This led to four possible diastereomers of the A ring, while the B ring, with four unknown stereocenters, has eight possible diastereomers—disregarding enantiomers.

Quantitative NOE calculations

When approaching quantitative NOE calculations of CM-A the fact that two slowly exchanging conformers were present could have been a considerable problem, as overlapping signals would be a sum of the two distinct conformational averages. The few overlapping signals present were thus excluded in the calculations.

Ring A

The stereochemistry of the A ring was apparent from the observations mentioned above and from simulated structures, due to differences in MAEs of the diastereomers and the rigidity for the ring. Selected distances from NOEs are presented in Table 3.6. The distances were selected as they are involved in NOEs and *J*-coupling

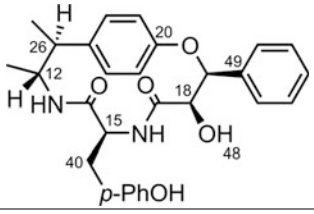
Table 3.6 Used distances and *J*-coupling constants for the determination of the stereochemistry of ring A (τ_m = 200 ms)

H1	H2	Dist. (Å)
12	24	2.57
13	15	2.18
13	22	3.10
13	26	2.37
15	16	3.13
15	41a	2.69
15	41b	2.80
16	18	3.26
16	41a	3.04
18	19	2.51
19	21	2.02
22	26	2.30
41a	41b	1.75[a]
H1	**H2**	***J* (Hz)**
12	13	9.9
12	26	9.3
15	16	9.2
15	40a	7.5
15	40b	7.9
18	19	0.0

[a]Reference distance

Table 3.7 MAEs from iteratively fitting back-calculated distances to experimental data, by minimizing MAE, for ring A to MM structures from simulations of diastereomers as indicated

Diastereomer	Distance (%)		*J*-coupling (Hz)	
	MAE	Max	MAE	Max
12*S*, 15*S*, 18*R*, 19*S*, 26*S*	3.2	7.6 (H12–H24)	0.4	1.1 (H15–H41a)
12*S*, 15*R*, 18*R*, 19*S*, 26*S*	6.2	18.5 (H16–H18)	1.0	1.7 (H18–H19)
12*S*, 15*S*, 18*S*, 19*S*, 26*S*	7.8	33.7 (H16–H18)	1.9	9.8 (H18–H19)
12*S*, 15*R*, 18*S*, 19*S*, 26*S*	6.5	21.4 (H19–H21)	1.6	5.2 (H18–H19)
12*R*, 15*S*, 18*R*, 19*S*, 26*R*	7.6	23.0 (H19–H21)	0.6	1.4 (H12–H13)



constants which include protons in the backbone of the ring. When these are compared to 3D structures the fits are as given in Table 3.7.

It is evident that the differences in distance were not that big between the different diastereomers of CM-A. While the change of stereochemistry at position 18 was problematic due to the increase in the possible *J*-coupling constant between H-18 and H-19, the differentiation of the diastereomer of 15 was more challenging. Still the fit of the data to the *L*-diastereomer was markedly better, especially if one identifies the distances which were not observed but should yield NOEs for the *D*-diastereomer. For example the distance of H-15 to H-18, which was 2.16 Å for the 15-*D* diastereomer, from the populations obtained by fitting the data, had no corresponding observed NOE in the spectra, which correlated well with the 15-*L* diastereomer (Fig. 3.9).

Once it was established that the ring was quite rigid, another approach to determine the configuration might be to assign stereocenters at random and use constraints from NOE and *J*-coupling constant data and floating chirality. This has been proposed by Baran et al. [22]. A similar approach could be used in the Schrödinger Suite using MacroModel. A couple of measures needed to be taken; the chirality should be allowed to change, a mix of torsional and low-mode sampling needed to be used, and the maximum low-mode steps needed to be quite big. In the current work, a minimum of 5 Å and a maximum of 25 Å was used (standard values are 3 Å and 6 Å, respectively). This allowed the molecule to easily switch chirality in the low mode steps, and due to the energy penalties on constraints the low energy output structures should primarily be the correct diastereomer. The distances in Table 3.6 were used as constraints with a penalty of 25 kJ/mol Å^2, and the *J*-coupling constants of H12–H13, H12–H26, H15–H16 and H18–H19 from Table 3.6 were translated to dihedral angles by $180 \pm n$ ($n = 180 \pm f$ (*J*)) or $0 \pm n$ ($n = 0 \pm f$ (*J*)) which is needed in MacroModel. The stereoisomer 12*S*,

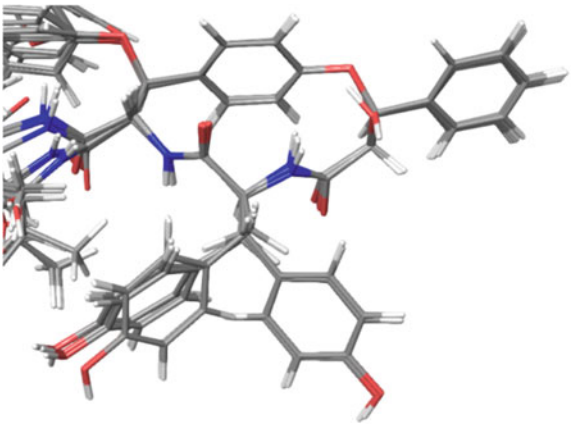


Fig. 3.9 The motif of ring A for the 12*S*, 15*S*, 18*R*, 19*S*, 26*S*-diastereomer (with the lowest MAE between experimental and theoretical data), determined from the fitting of NOE and *J*-coupling constant data to multiple 3D structures

Table 3.8 Result of floating chirality simulation, as described in the text, of ring A using distance and dihedral constraints from Table 3.6, using different energy cut-offs

Diastereomer	Population (%)		
	50 kJ/mol	25 kJ/mol	10 kJ/mol
12*S*, 15*S*, 18*R*, 19*S*, 26*S*	67.0	95.0	100.0
12*S*, 15*R*, 18*R*, 19*S*, 26*S*	32.7	5.0	0.0
12*S*, 15*S*, 18*S*, 19*S*, 26*S*	0.0	0.0	0.0
12*S*, 15*R*, 18*S*, 19*S*, 26*S*	0.3	0.0	0.0
12*R*, 15*S*, 18*R*, 19*S*, 26*R*	0.0	0.0	0.0

15*R*, 18*S*, 19*R*, 26*S* was chosen as the starting structure, and after the simulation the major diastereomer was 12*S*, 15*S*, 18*R*, 19*S*, 26*S* which constituted 67% of the structures determined in a window of 50 kJ/mol and 95% in a window of 25 kJ/mol from the determined energy minimum (3111 total structures), see Table 3.8 for more details.

Ring B

The B ring was more difficult to solve due to a higher degree of flexibility originating from the reduced rigidity, and a larger conformational space for the larger ring size. Since the diastereomers of ring A were solved, the three remaining chiral centers in the ring were varied and the best minimum MAE identified. The assignments of the diastereotopic protons of C-33 in the ring were allowed to interchange to best fit the data. Again, only select distances were used following the same selection criteria as for ring A, see Table 3.9. Also, distances between the protons of proline were ignored as these are shared between all possible conformers

Table 3.9 Used distances and *J*-coupling constants for the determination of the stereochemistry of ring B (τ_m = 200 ms)

H1	H2	Dist. (Å)
3	29	2.42
3	31	2.94
3	33b	2.42
4	6	2.34
4	29/31	3.16
4	33a	2.84
4	33b	3.76
6	9	2.06
6	37	2.73
9	10	2.71
9	37	2.53
10	12	2.22
10	13	2.96
10	38	2.98
26	30	2.07
29	33a	2.64
29	33b	2.55
31	33a	2.27
33a	33b[a]	1.75
37	38	2.56
H1	**H2**	***J* (Hz)**
3	4	8.8
3	33a	11.7
9	10	4.5
9	37	4.4

[a]Reference distance

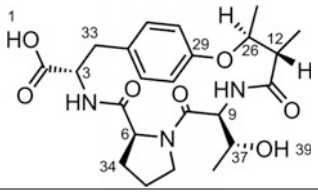
due to ring strain in the five-membered ring, and major resonance overlaps between protons in the ring.

The resulting MAEs from all diastereomers were almost equal (Table 3.10), and it was found more illustrious to compare the increase or decrease in correlation to the individual distances of the data. This is done in Table 3.11 as the difference between back-calculated and experimental distances (in %) and *J*-coupling constants (in Hz) for the relevant diastereomers. The all *S* diastereomer had the lowest MAE, and interchanging chiral centers led to several rather big errors in many distances and *J*-coupling constants between experiments and calculations. The two diastereomers 3*S*, 6*R*, 9*R*, 12*S*, 26*S* and 3*R*, 6*R*, 9*R*, 12*S*, 26*S* were the hardest diastereomers to dismiss by the data. The first one (3*S*, 6*R*, 9*R*, 12*S*, 26*S*) could be dismissed from the increase in the distance between H6 and H9 which translates to two alpha protons being too far apart in space. The latter diastereomer (3*R*, 6*R*, 9*R*, 12*S*, 26*S*) had an increase in the error of H26 distance to the aromatic ring in addition to an error in the H6–H9 distance.

A possible path to gain more certainty of the stereochemistry was to assume a better accuracy in the relative energies of the MMFFs force field, and limit the

Table 3.10 MAEs from iteratively fitting back-calculated distances to experimental data, by minimizing MAE, for ring B to MM structures from simulations of diastereomers as indicated

Diastereomer	Distance (%)		*J*-coupling (Hz)	
	MAE	Max	MAE	Max
3*S*, 6*S*, 9*S*, 12*S*, 26*S*	4.6	11.4 (H29–H33b)	0.0	0.1 (H9–H10)
3*R*, 6*S*, 9*S*, 12*S*, 26*S*	6.6	21.1 (H29–H33b)	1.2	1.2 (H3–H4)
3*S*, 6*R*, 9*S*, 12*S*, 26*S*	6.8	15.9 (H4–H6)	0.0	0.5 (H3–H33a)
3*S*, 6*S*, 9*R*, 12*S*, 26*S*	6.4	21.2 (H6–H9)	0.2	0.2 (H3–H4)
3*R*, 6*R*, 9*S*, 12*S*, 26*S*	8.2	25.2 (H3–H29)	0.5	0.5 (H3–H4)
3*R*, 6*S*, 9*R*, 12*S*, 26*S*	6.6	20 (H26–H30)	0.3	0.3 (H3–H4)
3*S*, 6*R*, 9*R*, 12*S*, 26*S*	6.0	17.9 (H6–H9)	0.8	0.8 (H3–H4)
3*R*, 6*R*, 9*R*, 12*S*, 26*S*	6.7	17.9 (H26–H30)	0.0	0 (–)



amount of structures based on energies. If the window was shrunk, from including structures within 50 kJ/mol of the identified minimum for each conformation, to 15 kJ/mol (which translates to a Boltzmann distribution population of ~0.2%), the stereochemistry is much more easily solved, as seen in Table 3.12.

The resulting structure consisted of all *L*-amino acids. To verify the assignment of *L*-amino acids, Marfey's analysis was performed and the proline was determine as *L*-proline [23, 24]. Proline was chosen, as acidic amino acids are less suitable for Marfey's analysis due to poor separation [23].

Full structure

The fitting of 3D structures to the full set of NMR data for the proposed diastereomer is illustrated in Tables 3.13, 3.14 and Fig. 3.10, and a representation of the structural space occupied (by multiple structures) is found in Appendix A4. The MAE was comparable to the MAE obtained from the much more rigid homomorphosin A, which was slightly surprising due to the more flexible nature of the compound. This could be a sign of overfitting, which is a problem when fitting multiple structures.

The stereochemistry of the novel NRP was thus solved based on NOEs and *J*-coupling constants. While the A ring is considered solved beyond doubt, the B ring proved more troublesome. The obtained solution yields by far the best correlation to the experimental data though, leaving it as the best possible solution. The problem would not have been properly addressed using a single structure due to the degree of flexibility and resulting high differences of data to 3D structures for all diastereomers leading to poor discrimination of their consistency with the NMR data.

Table 3.11 Error in distances/*J*-coupling constants in %/Hz compared to the 3*S*, 6*S*, 9*S*, 12*S*, 26*S*-diastereomer

Change in distance errors (%)								
H1	H2	3*R*	6*R*	9*R*	3*R*6*R*	3*R*9*R*	6*R*9*R*	3*R*6*R*9*R*
3	29	**10.2**	−1.9	**8.5**	**18.6**	**6.0**	−0.8	**5.0**
3	31	2.3	−0.4	3.0	**9.7**	−0.5	−0.1	1.7
3	33b	−1.7	1.1	−2.3	−2.5	−1.0	−1.9	−0.2
4	6	**6.7**	**14.4**	2.9	0.6	**11.0**	4.8	1.4
4	33a	5.1	4.2	4.3	4.8	2.4	1.8	3.4
4	33b	−1.9	−1.8	−1.2	−0.4	−0.9	−0.8	−1.2
6	9	−0.9	**6.3**	**11.7**	**13.3**	2.7	**8.4**	**6.4**
9	10	−1.2	−1.2	0.5	−2.5	−1.3	0.0	0.1
9	37	1.7	0.1	1.7	0.6	3.4	**5.7**	4.5
10	12	−1.0	2.0	−0.3	4.7	1.5	3.5	3.2
10	13	1.4	**15.1**	3.8	0.9	2.4	−0.3	2.1
10	38	−2.5	1.7	2.2	4.4	−4.5	−2.4	0.3
26	30	**5.9**	1.2	−0.9	**12.5**	**13.7**	4.4	**11.5**
29	33a	*−6.3*	1.5	0.6	−4.9	−3.8	1.6	*−5.3*
29	33b	**9.7**	−2.0	−3.2	−0.8	0.4	0.0	1.2
31	33a	**9.1**	0.5	0.7	**5.7**	**5.0**	1.2	3.2
37	38	0.0	−0.1	0.2	0.1	0.2	−0.2	0.1
Change in J-coupling constants (*Hz*)								
H1	H2	3*R*	6*R*	9*R*	3*R*6*R*	3*R*9*R*	6*R*9*R*	3*R*6*R*9*R*
3	4	**1.2**	0.0	0.2	**0.5**	0.3	**0.8**	0.0
3	33a	0.0	**0.5**	0.0	0.0	0.1	0.0	0.0
9	10	0.0	−0.1	−0.1	−0.1	−0.1	−0.1	−0.1
9	37	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Diastereomers varied as indicated. Bold/Italic indicates an increase/decrease in error of more than 5%/0.5 Hz

Table 3.12 Fit, obtained by iteratively minimizing MAE, of NOE derived distances for ring B to those of MM structures from simulations of diastereomers

Diastereomer	Distance (%)		*J*-coupling (Hz)	
	MAE	Max	MAE	Max
3*S*, 6*S*, 9*S*, 12*S*, 26*S*	5.0	14.2 (H10–H38)	0.1	0.3 (H9–H10)
3*R*, 6*S*, 9*S*, 12*S*, 26*S*	11.4	38.4 (H10–H13)	1.1	3.3 (H9–H10)
3*S*, 6*R*, 9*S*, 12*S*, 26*S*	16.3	120.2 (H6–H9)	1.0	3.6 (H9–H10)
3*S*, 6*S*, 9*R*, 12*S*, 26*S*	16.7	122.1 (H6–H9)	0.9	3.4 (H9–H10)
3*R*, 6*R*, 9*S*, 12*S*, 26*S*	16.0	121.6 (H6–H9)	0.9	3.6 (H9–H10)
3*R*, 6*S*, 9*R*, 12*S*, 26*S*	14.5	120.4 (H6–H9)	0.9	2.6 (H9–H10)
3*S*, 6*R*, 9*R*, 12*S*, 26*S*	8.7	20.8 (H6–H9)	1.2	2.5 (H9–H10)
3*R*, 6*R*, 9*R*, 12*S*, 26*S*	8.5	30.2 (H26–H30)	0.2	0.7 (H3–H4)

Structures within 30 kJ/mol of the global minimum used

Table 3.13 Comparison of the distances from NOE intensities and 3D structures using multiple structures from MM (τ_m = 200 ms)

H1	H2	Exp. dist.	3D Dist.	Diff (Å)	Diff (%)
3	29	2.52	2.52	0.00	0.2
3	31	3.06	3.07	0.01	0.4
3	33b	2.52	2.51	0.01	0.5
4	6	2.43	2.34	0.09	3.8
4	33a	2.96	2.83	0.14	4.6
4	33b	3.91	3.85	0.06	1.5
4	35b	3.29	3.08	0.22	6.5
4	36b	3.46	3.47	0.02	0.5
6	9	2.15	2.31	0.16	7.3
6	34a	2.39	2.35	0.04	1.7
6	34b	2.59	2.73	0.14	5.5
9	10	2.82	2.80	0.01	0.5
9	37	2.64	2.69	0.05	1.8
10	12	2.31	2.17	0.14	6.1
10	13	3.08	3.01	0.07	2.2
10	38	3.11	3.32	0.21	6.7
12	24	2.57	2.75	0.19	7.3
13	15	2.18	2.13	0.05	2.5
13	22	3.10	3.23	0.13	4.2
13	26	2.75	2.92	0.17	6.0
15	16	3.13	2.99	0.15	4.6
15	40a	2.66	2.73	0.06	2.3
15	40b	2.80	2.76	0.05	1.6
16	18	3.26	3.29	0.02	0.6
16	40a	3.04	2.86	0.18	5.8
18	19	2.51	2.64	0.50	5.2
19	21	2.02	2.14	0.13	5.9
22	26	2.30	2.32	0.12	0.9
26	30	2.16	2.19	0.02	1.2
29	33a	2.75	2.92	0.03	6.3
29	33b	2.65	2.80	0.17	5.6
31	33a	2.36	2.48	0.15	5.0
33a	33b	1.82	1.75	0.12	4.0
34b	36b	2.40	2.43	0.07	1.3
35b	36a	2.30	2.42	0.03	4.9
37	38	2.67	2.62	0.11	1.7
38	43/45	3.60	3.64	0.05	1.1
40a	40b	1.75[a]	1.75	0.05	0.0
			MAE	0.10	3.4

All distances are in Å. [a]Reference distance

Table 3.14 Comparison of experimental *J*-coupling constants and calculated constants by the Karplus or HLA equation in Hz, from 3D structures using multiple structures from MM

H1	H2	*J* exp.	*J* calc.	Diff
3	4	8.8	8.6	0.2
3	33a	11.7	11.8	0.1
9	10	4.5	4.8	0.3
9	37	4.4	4.4	0.0
12	13	9.9	9.4	0.5
12	26	9.3	9.7	0.4
15	16	9.2	9.1	0.0
15	40a	7.5	7.4	0.0
15	40b	7.9	7.9	0.0
18	19	0	0.8	0.8

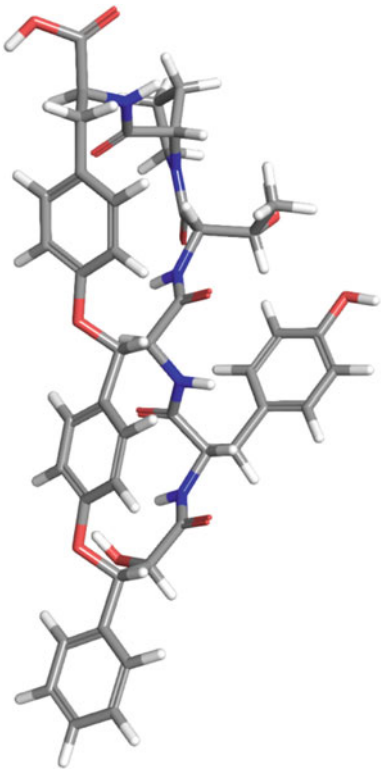


Fig. 3.10 Highest populated structure from the NOE and *J*-coupling constant data. Used as a representative structure for the structural space inhabited by cyclomorphosin A, suggested from NOE and *J*-coupling constant data

Due to a very limited amount of compound, a similar analysis of CM-B was not possible, and not even all ^{13}C resonances were observed in the spectra. The assignment is thus based on the very similar chemical shifts, spin systems and NOEs at long mixing times. The A ring is deemed of identical, or fully reversed, stereochemistry due to very similar *J*-coupling constants and chemical shifts, while the stereochemistry of the B ring was not solved due to differences in the amino

acids and an increased flexibility from substituting a proline to a threonine. The minor conformer observed in the CM-B spectra was in the ratio 10:1, which is contributed to relieved strain in the B ring from removal of the proline, and it is suggested that the minor conformation is based in changes in the B ring for both structures.

3.1.5 Homomorphosterol

The novel sterol homomorphosterol, see Fig. 3.11, was isolated from *A. homomorphus*. In the work of the Master Thesis "Chemistry of Black Aspergilli" the stereochemistry was wrongfully assigned, and the determination of the right diastereomer is thus included here. This is a good case of usage of quantum chemical calculations in solving the stereochemistry of rigid compounds.

As indicated in Fig. 3.11 large ^{5}J-coupling constants of 6 and 5 Hz were observed which have also previously been reported for various sterols [25]. This feature should be possible to address with the usage of computational chemistry. Featured important HMBC correlations and NOEs are depicted in Fig. 3.12, and while the stereochemistry is in principle solvable from qualitative NOE analysis more certainty was desired. It should be noted that only data with a mixing time of 800 ms were available, and with no build-up curve and probably a too long mixing

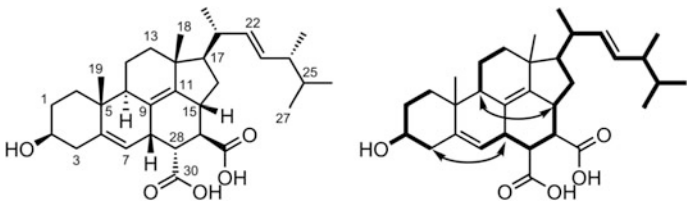


Fig. 3.11 The structure of homomorphosterol (left) and the spin systems as determined by DQF-COSY and HSQC-TOCSY (right). Arrows indicate observed ^{5}J-coupling constants

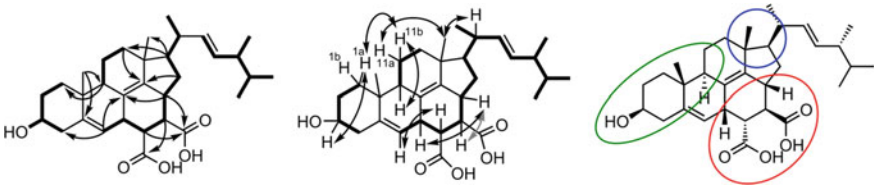


Fig. 3.12 Selected HMBC connectivities (left) and NOEs (center) of homomorphosterol. The grey arrow indicates that scalar coupling induced artifacts in the NOESY spectra made it impossible to judge if an NOE was present. Right: The three stereo-clusters in the cyclic part of homomorphosterol. Stereocenters within the green and blue circles are thought equal to the stereochemistry of ergosterol

time, quantitative NOE analysis was not pursued due to a probable, ultimately unknown, degree of spin diffusion.

An approach to solve the stereochemistry of rigid compounds is by DFT calculation of chemical shifts and *J*-coupling constants. The utilization of chemical shifts and a probability function was deemed useful in this case [26]. It should be noted that homomorphosterol features three clusters of stereocenters as indicated in Fig. 3.12. It was proposed that the biosynthesis of homomorphosterol involved a Diels-Alder like reaction from an oxidized ergosterol and two of the clusters, blue and green in Fig. 3.12, are structurally most likely identical to the structure of ergosterol [27]. A similar compound was previously reported where a similar biosynthesis as the one suggested in Fig. 3.13 was proposed, utilizing maleimide instead of maleic or fumaric acid [28].

This leaves essentially three stereocenters to be solved; H-3 and H-6 (which are coupled), H-30 and H-31. DFT-optimized structures of truncated diastereomers, with the flexible side-chain at C-17 changed to a methyl group as given in Fig. 3.14, were used in a DP4 probability analysis, as calculated by Eq. (3.1) [26].

$$P(i|\delta_N) = \frac{\prod_{k=1}^{N} 1 - T^v\left(\left|\delta_{scaled,k}^{-i} - \delta_{exp,k} - \mu\right|/\sigma\right)}{\sum_{j=1}^{m}\prod_{k=1}^{N} 1 - T^v\left(\left|\delta_{scaled,k}^{-i} - \delta_{exp,k} - \mu\right|/\sigma\right)} \tag{3.1}$$

where $\mu = 0$, $v = 11.38$ (^{13}C) or 14.18 (^{1}H) and $\sigma = 2.306$ (^{13}C) or 0.185 (^{1}H) ppm.

Only a single conformer was identified for each diastereomer due to the highly rigid nature of the compound. The rigidity also results in rather big differences in the spatial structure for each diastereomer as illustrated in Fig. 3.14, which accordingly yields differences in theoretical chemical shifts and *J*-coupling constants.

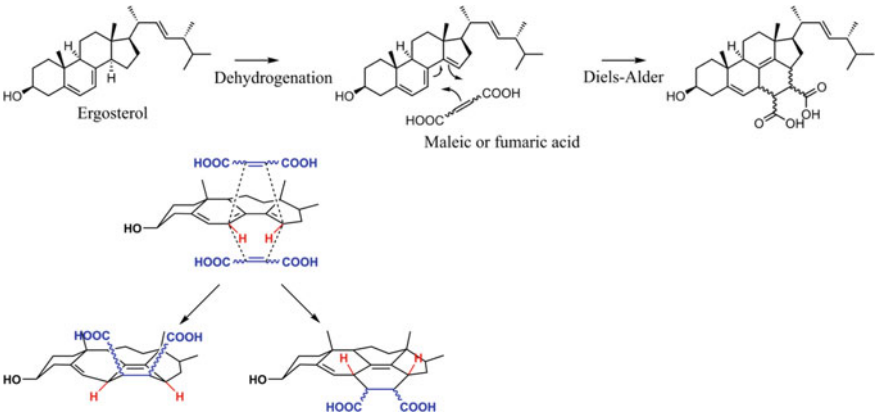


Fig. 3.13 Top: The suggested bio-synthetic pathway to homomorphosterol. First a proton is removed from ergosterol to give a conjugated system, and the resulting triene reacts with either maleic or fumaric acid in a Diels-Alder-like reaction to gain the final structure. Left: Resulting stereochemistry of a Diels-Alder reaction with the triene

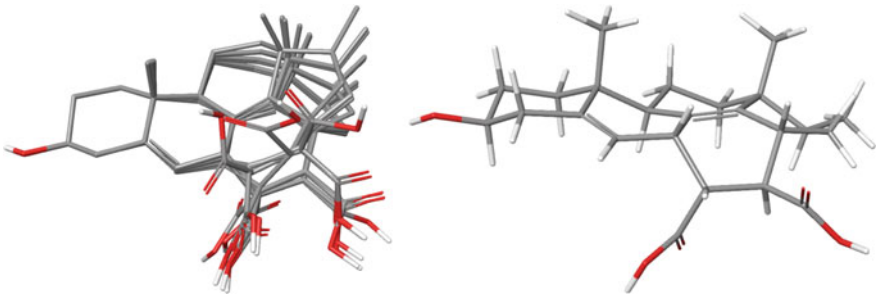


Fig. 3.14 Left: Overlay of the eight stereoisomers. A quite big difference in the occupied structural space is observed. Right: The stereoisomer (*RRRS*) which showed the best correlation to the available data. All structures are truncated at C-18 to focus on the rigid core

The chemical shifts are collected in Appendix A4. The comparison of the DP4 probability of experimental versus theoretical chemical shifts of the diastereomers is given in Table 3.15, which clearly shows that the (*RRRS*) structure was favored. Especially the proton chemical shifts were discriminative which was surprising since the carbon chemical shifts hold a larger chemical shift range and thus could potentially be more discerning. This may be due to the added electronic complexity when modelling the ^{13}C nuclei compared to the much simpler protons. The MAE for the (*RRRS*) diastereomer was 0.12 ppm for ^{1}H and 1.68 ppm for ^{13}C chemical shifts which resembles the 20 times larger chemical shift range of ^{13}C and might explain why the proton chemical shifts were more discerning. This is in good agreement with findings from the literature [26, 29, 30].

Also the *J*-coupling constants, including the large long-range constants, were investigated and gave the same conclusion for the stereochemistry, see Table 3.16. As the (*RRRS*) structure did not violate any qualitative NOE correlations, this structure was assumed to be valid. For more discussion on the DFT calculation of *J*-coupling constants see Sect. 4.2.1.

In conclusion the stereochemistry of a novel steroid was determined from calculations of NMR properties, based on a priori knowledge of the dominating biosynthetic pathway of steroids in fungi, and qualitative NOEs. Note that the a priori knowledge can lead to the absolute stereochemistry if the assumptions are correct. If incorrect, the relative stereochemistry is still upheld from qualitative

Table 3.15 DP4 probability analysis of stereoisomers of homomorphosterol (values in %)

	(*RRSS*)	(*RRRS*)	(*RSRS*)	(*RSSS*)	(*SRSR*)	(*SRRR*)	(*SSSR*)	(*SSRR*)
DP4$^{13}C/^{1}H$	**0**	*99.9*	0.1	**0**	**0**	**0**	**0**	**0**
DP4^{13}C	0.1	*58.8*	11.5	8.1	0.7	3.9	16.8	0
DP4^{1}H	**0**	*99.7*	0.3	**0**	**0**	**0**	**0**	**0**

The first row covers ^{13}C and ^{1}H chemical shifts. The best and worst fit to the data is marked in bold and italic respectively. Shielding tensors calculated by MPW1PW91/6-311 + G(2d,p) using GIAO from B3LYP/6-31 g(d) optimized structures

Table 3.16 Experimental and theoretical *J*-coupling constants in Hz for stereoisomers of homomorphosterol

J (Hz)	Meas.	(*RRSS*)	(*RRRS*)	(*RSRS*)	(*RSSS*)	(*SRSR*)	(*SRRR*)	(*SSSR*)	(*SSRR*)
H3–H31	9.9	9.7 *0.4*	10.0 ***0.1***	10.9 *0.8*	9.8 *0.3*	11.5 *1.4*	9.2 *0.9*	11.3 *1.2*	4.2 *5.9*
H6–H30	6.9	4.2 *2.7*	7.7 ***0.8***	11.0 *4.1*	10.3 *3.4*	10.8 *3.9*	10.8 *3.9*	8.5 *1.6*	10.8 *3.9*
H30–H31	5.7	4.4 *1.3*	5.3 ***0.4***	11.4 *5.7*	12.0 *6.3*	11.5 *5.8*	7.4 *1.7*	6.2 *0.5*	4.3 *1.4*
H6–H9a*	6.0	6.9 *0.9*	6.1 ***0.1***	6.1 *0.1*	7.0 *1.0*	3.5 *2.5*	3.5 *2.5*	4.0 *2.0*	4.9 *1.1*
H3–H14*	5.0	5.2 *0.2*	4.6 *0.4*	4.8 *0.2*	5.1 ***0.1***	5.2 *0.2*	5.4 *0.4*	5.3 *0.3*	4.3 *0.7*

Indented numbers are the absolute difference compared to the measured couplings. Coupling constants were measured from a 1D ^{1}H spectrum or DQF-COSY(*). The best and worst fits to the data are marked in bold italic and underline, respectively. Theoretical coupling constants calculated by B3LYP/6-31 g(d, p) u + 1 s from B3LYP/6-31 g(d) optimized structures

NOEs and the calculations above. The rigid structure was here a necessity to avoid too many structures per stereoisomer, though averaging is possible [29]. This is a prime example of how DFT computation of NMR observables can be a crucial method in solving unknown structures, a method which will only find increased usage with the introduction of more observables, e.g. in the form of more long-range coupling constants as described in Chap. 4.

3.2 Other Natural Products

In addition to the compounds isolated from *A. homomorphus*, other natural compounds were investigated. Some of these are included briefly in the following.

3.2.1 Aculenes

The aculenes (Fig. 3.15), as well as the epi-10,23-dihydro-24,25-dehydroaflavine discussed in the next section, are small molecules isolated by Dr. Lene M. Pedersen, for which the 3D structure were determined from *J*-coupling constants, structural calculations and select qualitative NOEs [31]. The compounds were isolated from another black aspergillus; the fungus *A. aculeatus* [10, 11].

The stereochemistry of the aculenes was determined mostly from *J*-coupling constants. It was evident that the two stereocenters in the 7-membered ring of aculene A needed to be (*R*)/(*S*) from theoretical *J*-coupling constants, see Table 3.17. This is independent of the assignment of the diastereotopic protons at

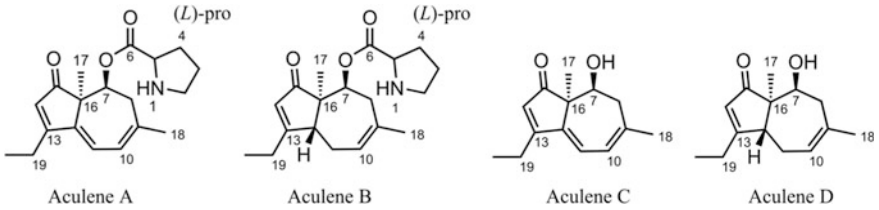


Fig. 3.15 The structures of aculene A-D with the determined stereochemistry shown

Table 3.17 Measured and theoretical *J*-coupling constants in Hz for the possible stereoisomers of the aculene A. Theoretical coupling constants calculated by B3LYP/6-31 g(d,p) u + 1 s from B3LYP/6-31 g(d) optimized structures

H1	H2	Exp.	(RS) or (SR)		(RR) or (SS)	
			Major	Minor	Major	Minor
7	8	4.4	4.5	11.0	9.3	11.3
7	8′	2.8	2.6	5.2	0.6	5.9

C-8, as an exchange of conformation would not improve the fit to the data of any stereoisomer. From the *J*-coupling constants it was suggested that the minor conformer was lowly populated. The stereochemistry of the proline was solved by Marfey's reagent, but as it may rotate freely around the C7–O bond, it was not possible to correlate the stereocenter to the remaining structure.

Due to a quite plausible biosynthetic relationship of the aculene A, B and C, and similar *J*-coupling constants and NOEs, the stereochemistry was thought to be identical. The remaining stereocenter of aculene B was determined qualitatively from NOEs (Fig. 3.16). It should be noted that aculene D was not isolated but suggested from HRMS data, and thus no NMR data were available in support.

3.2.2 Epi-10,23-Dihydro-24,25-Dehydroaflavine

The stereochemistry of epi-10,23-dihydro-24,25-dehydroaflavinine was solved by the use of NOEs and *J*-couplings. The relative stereochemistry of epi-10,23-dihydro-24,25-dehydroaflavinine was established by performing conformational searches on different diastereomers and by comparing observed distances to the back-calculated distances from ISPA, and by comparison of experimental ^{3}J-couplings to HLA back-calculations. The relative stereochemistry was determined to be equal to that of the original 10,23-dihydro-24,25-dehydroaflavinine [32, 33], but as the optical rotation was of the opposite sign, compared to the literature, the enantiomer was most likely isolated. The structure that resulted in the lowest MAE was further optimized by HF and DFT to give the reported structure.

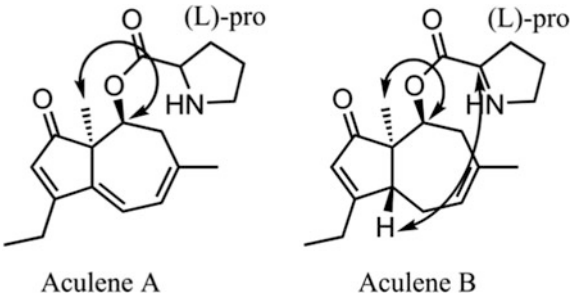


Fig. 3.16 Qualitative NOEs used for solving the structures of aculene A and B (τ_m = 800 ms)

Again a lower error was obtained if the rotating groups, the isopropylene and the alcohol, were allowed to rotate, see Tables 3.19 and 3.20, as the inclusion of multiple structures eliminated most of the observed errors when using only a single structure. It should be noted that while the single structure was optimized to a DFT level, the multiple structures are used directly from the force field simulations, accounting for some of the differences in the fit of the rigid parts and in the reference distance.

By using the ISPA approach it was possible to unambiguously assign the diastereomer of the compound. This might have been possible using qualitative data and a model, which was how the enantiomer was previously solved [33], but the complicated geometry of the structure made ISPA an easier approach, where one did not have to guess whether the stereochemistry was actually the best possibility. While the alcohol and vinyl groups were rotating in the identified conformer populations, the orientation of the indole ring was fixed, in good correlation to the force field energies, where a reorientation of the indole ring of 180° gave rise to an approximate rise in energy of 16 kJ/mol (a population of ~0.16%). Even with the discriminative properties of NOEs this population should not be detected. While the force field energies may be prone to errors, the obtained population of fitting the rotamers correlated pretty well to that of the energies, as seen in Table 3.18.

In order to confirm the assignment of the diastereomer a constrained conformer search with floating chirality, using only data between methine and methyl protons was also conducted, following the procedure given for cyclomorphosin A, which resulted in only the given diastereomer in a 50 kJ/mol window (Fig. 3.17 and 3.18).

Table 3.18 Population of the minimum energy conformation (1) and different rotamers of C-23 (2), C-19 (3) and both (4). Populations from force field energies and NOE data

Conformer	Rel. E (kJ/mol)	Rel. FF pop. (%)	Rel. NOE pop. (%)
1	0	46	56
2	1.26	28	12
3	2.65	16	23
4	3.89	10	9

Table 3.19 Comparison of the distances from NOE intensities and from 3D structures using a single DFT optimized structure (B3LYP/6-31 g(d)) or multiple structures from MM (τ_m = 150 ms)

		Single			Mult		
H1	H2	Exp. Dist.	Calc. Dist.	%	Exp. Dist.	Calc. Dist.	%
1	2	2.64	2.50	5.6	2.62	2.48	5.5
1	5	3.03	2.83	7.1	3.01	2.81	6.7
5	18	2.72	2.40	13.3	2.70	2.46	8.8
5	11	2.41	2.37	1.7	2.40	2.62	9.3
5	10	2.54	2.67	4.9	2.53	2.56	1.5
2	27	2.89	2.96	2.4	2.88	2.99	3.5
2	23	2.58	2.46	4.9	2.56	2.56	0.1
2	25	3.01	2.96	1.7	3.00	3.04	1.3
25	23	2.39	2.31	3.5	2.37	2.39	0.9
25	10	2.78	4.58	39.3	2.77	2.82	2.0
25	25′	1.85	1.85	0.0	1.84	1.85	0.6
19	26	3.57	3.53	1.0	3.56	3.64	2.3
19	18′	2.33	2.43	3.3	2.33	2.45	5.0
19	22	2.02	2.14	5.6	2.01	2.23	10.7
19	18	2.32	2.41	3.7	2.31	2.44	5.5
19	11	3.37	2.83	19.1	3.35	2.95	12
19	10	1.95	1.88	3.7	1.93	2.01	3.8
30	29	3.45	3.39	1.8	3.44	3.65	6.0
30	18′	2.34	3.32	25.3	2.33	2.41	3.5
30	22	2.68	2.25	19.1	2.66	2.70	1.3
30	21	2.65	2.06	28.6	2.64	2.54	3.6
10	26	2.28	2.33	2.1	2.72	2.86	5.4
10	11	2.24	2.32	3.4	2.23	2.33	4.7
23	27	2.06	2.05	0.5	2.40	2.49	3.7
11	12	2.12	2.24	5.4	2.18	2.28	4.5
11	13	2.70	2.61	3.4	2.68	2.67	0.3
11	16	1.86	2.21	15.8	1.85	2.00	7.7
16	28	2.68	2.84	5.9	2.67	2.60	2.6
21	29	2.46	2.61	6.0	2.45	2.64	7.6
16	17′	2.10	2.44	13.9	2.09	2.16	3.4
16	13	1.93	2.02	4.5	1.91	2.09	9.0
22	22′	1.75	1.74	0.6	1.74	1.74	0.0
17	29	2.77	2.79	0.9	2.76	2.84	3.0
17	17′	1.74	1.75	0.6	1.71	1.75	2.3
21′	27	2.58	2.52	2.4	2.57	2.62	1.7
13	13′	1.75[a]	1.75	0.0	1.74[a]	1.74	0.0

(continued)

Table 3.19 (continued)

		Single			Mult		
H1	H2	Exp. Dist.	Calc. Dist.	%	Exp. Dist.	Calc. Dist.	%
17′	28	2.82	2.95	4.5	2.81	2.87	2.1
14′	28	2.54	2.60	2.3	2.53	2.70	6.8
25′	26	2.85	2.92	2.5	2.84	2.88	1.2
	MAE			6.9			4.1

All distances in Å. [a]Reference

Table 3.20 Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation or DFT calculation from a single DFT optimized 3D structure, in Hz

H1	H2	Exp.	HLA	DFT
10	23	13.3	12.7	12.4
10	11	5.0	4.3	5.1
21′	22	13.5	12.6	11.9
22	23	13.5	11.9	13.7
16	17	11.3	12.5	13.7
17	18	11.3	13.4	14.1
21′	22′	3.4	4.8	4.4
21	22	4.3	4.5	4.4
22’	23	5.6	4.8	5.3

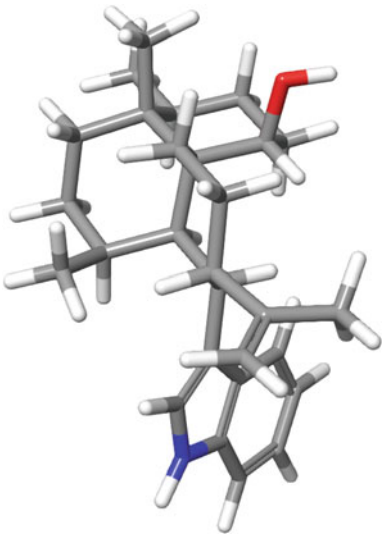


Fig. 3.17 3D structure of epi-10,23-dihydro-24,25-dehydroaflavine with the best correlation to the NMR data

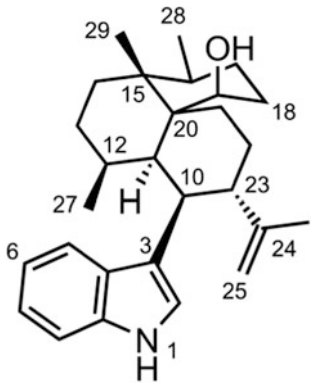


Fig. 3.18 Numbered structure of epi-10,23-dihydro-24,25-dehydroaflavine

3.3 Synthetic Peptides

The experimental parts of this section are based on two articles [34, 35]. More space is used discussing the results from NMR spectroscopy than the various biological roles and otherwise exiting potential of the included compounds, which will only be briefly touched upon. For a more in depth description the reader is referred to the articles [34, 35] and the literature referred to in the text.

3.3.1 Azumamides

The azumamides are natural products, originally isolated from the marine sponge *Mycale izuensis* [36]. The compounds are cyclic non-ribosomal peptides (NRPs), illustrated in Fig. 3.19, which were shown to inhibit histone deacetylase (HDAC) activity [36].

HDACs are a group of enzymes with the common trait that they deacetylate the *N*-*ε*-position of lysine residues at the *N*-terminal tail of histones [37, 38]. In groups of eight, histones proteins form the nucleosome; the repeating unit in chromatin, which is the highly ordered structures that store DNA in eukaryotic cells [37, 38]. Gene-transcription is regulated by interactions between the *N*-terminal end of the histones and DNA, and the interactions are modulated by e.g. the degree of histone acetylation, methylation or phophorylation [37, 38]. Histone acetylation will therefore affect gene transcription as the degree of histone acetylation correlates positively with transcription [38], and a decrease in acetylation has been linked to cancer due to a lower expression of regulatory genes [37]. Compounds that inhibit HDAC activity may thus be used as anti-cancer agents, spurring the interest in understanding and improving the activity of the compounds. Humans have 18 HDACs, of which 11 are Zn^{2+} dependent and were the focus of the study.

	R_1	R_2	R_3
A:	NH_2	H	Me
B:	NH_2	OH	Me
C:	OH	OH	Me
D:	NH_2	H	H
E:	OH	H	Me

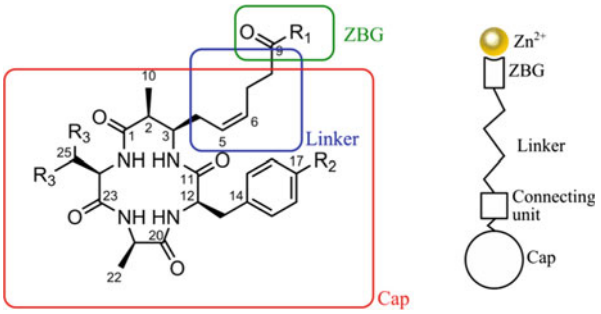


Fig. 3.19 The structure of the natural azumamides A-E with indicated pharmacophore, and a schematic pharmacophore of HDAC inhibitors [36, 37]

The mode of action of most HDAC inhibitors is by interacting with Zn^{2+} in the binding site of the HDAC enzymes, effectively blocking the active site. HDAC inhibitors are structurally quite different compounds but generally have a pharmacophore as indicated in Fig. 3.19 [37].

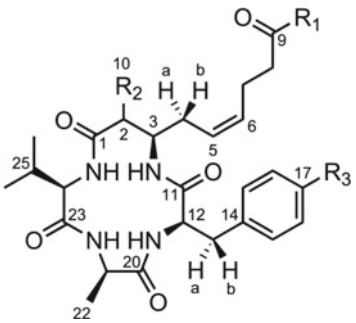
The group of Professor Christian A. Olsen (CAO), then DTU now University of Copenhagen, were interested in investigating analogues of the natural azumamides, synthesized in lab from total synthesis, to develop novel anti-cancer compounds, and it was in that regard desired to correlate structural features to HDAC inhibition. Thus an NMR study of natural and unnatural analogues of azumamides was conducted in collaboration with Dr. Alex Maolanon and Dr. Jesper Villadsen both from the group of CAO, who synthesized the investigated molecules, and Dr. Niels Christensen from the group of Associate Professor Peter Fristrup, who performed the calculations and theoretical docking studies. For more information regarding the comprehensive synthetic work, activity testing and docking results the reader is referred to the article [34].

Prior work

A lot of work has been put towards total synthesis and HDAC activity of the native azumamides as well as structural analogues thereof [39, 40]. This, often synthetic or biological work, is mostly outside the scope of this thesis and it was chosen to keep the focus relevant to the presented work – namely 3D structural information from NMR spectroscopy. The 3D structures of azumamide E [39, 40] and an unnatural analogue 2*S*, 3*R* azumamide E [40] have previously been published from constrained MD calculations. The comparable structures of azumamide E are alike; the amide protons NH-1,3 and 4 point upwards from the plane of the ring while NH-2 points slightly downwards and the phenyl ring is superimposing the ring akin to a lid. These structures are used for comparison in the following sections.

Investigated structures

Three structures were chosen for NMR investigation; the natural azumamide A (azu A) and two unnatural analogues desmethyl-azumamide C and epimethyl-azumamide E (des D and epi E). The structures are given in Fig. 3.20.



	R_1	R_2	R_3
Azu A:	NH_2	(*S*)-Me	H
Des C:	OH	H	OH
Epi E:	OH	(*R*)-Me	H

Fig. 3.20 Structure of the investigated structural analogues azumamide A (azu A), des-azumamide C (des C) and epi-azumamide E (epi E). Stereochemistry, including the diastereotopic protons important for the 3D structural investigation, is indicated

The structural analogues exhibited in vitro activity in the order azumamides > des-methyl azumamides ≫ epi-methyl azumamides [34]. It was proposed that conformational differences in the ring conformation could be the reason for the differences in activity. The slight differences in the groups R_1–R_3 (Fig. 3.20) were deemed of lesser importance for the ring conformation, and the general conformational spaces occupied by the structures. Unfortunately the in vitro activity was not retained in in vivo studies of a human Burkitt's lymphoma cancer cell line.

As the interesting properties of the azumamides are all occurring in an aqueous environment, it would be beneficial to obtain data in water. Attempts to dissolve the compounds in water and in a water/DMSO-d_6 mixture with up to 15% DMSO-d_6 resulted in very low *S*/*N* ratios, and spectra acquired in pure DMSO-d_6 were used for further investigations. It is acknowledged that this may lead to errors in the structure compared to that of a biological environment, but it could not be averted.

Assignment of diastereotopic protons 4 and 13

The diastereotopic protons at C-4 and C-13 were not easily assigned due to possible free rotation around the bonds C3–C4 and C12–C13. It was realized from NOE-derived distances and *J*-coupling constants, that free rotation was generally not present, due to differences in NOE intensities and the size of *J*-coupling constants to the diastereotopic protons, allowing assignment of the protons. Only the rotation around C3–C4 of the des-methyl variant was observed to be unrestricted. The protons were thus assigned by minimizing the error of NOEs and *J*-coupling constants. This might not lead to the right answer, but without an assignment a lot of data and information was lost, and both data types suggest the given assignments unambiguously, making it the most viable option.

The assignment results in structures that fit the NMR data displaying the aromatic ring of Phe or Tyr outside the cylinder of the ring, in contrast to previously published structures, which is in good correlation with docking studies, *vide infra*.

Constrained versus unconstrained calculations

Constrained MD using experimental observables such as relative NOE intensities and *J*-coupling constants has been a favored method of obtaining 3D structures [39, 40]. While the approach has validity for macromolecules, where the amount of flexibility is beyond what even modern computational chemistry can reasonably be expected to handle, for relatively ordered systems a lot of information will inevitable be lost. This is in part due to the constrains being applied as distances and dihedral angles, while the observables are in fact NOEs and *J*-coupling constants, which will lead to a faulty averaging of the data, see Eqs. (3.2)–(3.5). Here it should be noted that not all structural simulation programs are capable of performing an averaging routine; it is obvious that information is lost if averaging is disregarded. The distinction in (3.2)–(3.5) may seem of less importance but will lead to large error in conformational populations.

$$\eta_{obs,i} = k r_i^{-6} \quad \text{for } i = 1 \tag{3.2}$$

$$\eta_{obs,i} = k \langle r_i^{-6} \rangle \quad \text{for } i > 1 \quad \text{not} \quad \eta_{obs,i} = k \langle r_i \rangle^{-6} \tag{3.3}$$

$$J_{obs,i} = f(\theta_i) \quad \text{for } i = 1 \tag{3.4}$$

$$J_{obs,i} = \langle f(\theta_i) \rangle \quad \text{for } i > 1 \quad \text{not} \quad J_{obs,i} = f(\langle \theta_i \rangle) \tag{3.5}$$

The other possibility is to use unconstrained simulations and structural averaging of conformers. This is an approach used earlier in this thesis that is gaining popularity and being applied in a slightly altered form also in macromolecular NMR [41]. To achieve this, a couple of requirements are apparent. First the simulations need to cover the structural space that the actual in vitro compounds occupy. But if too much space is covered it may hamper the fitting of the experimental data to the 3D structures. This means that while an amount of conformational flexibility is a good thing, too much flexibility may lead to unsolvable systems. Also one has to accept a couple of approximations; that the correlation time is independent of conformation and that the averaging motions are of a timescale that allows the usage of r^{-6} averaging. For the azumamides, due to an at least somewhat restricted conformational space due to the cyclic system, it should be possible to use unconstrained simulations and increase the coverage of the conformational space and thus obtainable molecular knowledge. For comparison both strategies were tried.

Constrained structure

Constraints in the form of NOE distances (Table 3.21) were applied to a structure of azumamide A, and the structure was optimized, depicted in Fig. 3.21. This structure was in good correlation to the published structures of azumamide E from constrained optimizations, with the exception of the orientation of the aromatic ring as described earlier, and the extension of the linker chain, which was generally less ordered, and was not considered beyond the double-bond [39, 40].

Table 3.21 Experimental and back-calculated distances in Å, including differences in %, and *J*-coupling constants in Hz for the azumamide analogues (τ_m = 150 ms)

Azumamide A					Des-azumamide C					Epi-azumamide E				
H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%
N1	N2	2.13	2.15	1.3	N1	N3	2.79	2.8	0.4	N1	N2	1.88	1.91	1.1
N1	N4	3.11	3.11	n	N1	2b	2.29	2.49	7.8	N1	2	2.5	2.49	0.3
N1	3	2.62	2.66	1.4	N1	3	2.73	2.7	1	N4	2	2.15	2.12	1.1
N1	4a	3.12	3.12	0	N1	4a/4b	3.11	3	3.6	N2	13a	2.92	2.92	0.1
N1	4b	3.06	3.06	0	N1	12	2.98	3.11	4.5	N2	22	3.12	3.13	0.3
N1	10	2.75	2.8	1.8	N1	13a	2.79	2.72	2.3	N3	N4	2.39	2.29	4.2
N1	13a	2.49	2.49	0	N1	13b	2.99	3	0.6	N4	25	2.98	2.99	0.1
N1	13b	2.42	2.44	0.7	N1	N2	1.99	1.91	4.4	N4	26/27	2.93	2.75	6.6
N2	13a	2.64	2.64	0	N2	N3	2.21	2.2	0	2	3[a]	3	3	0
N2	13b	2.89	2.89	0	N2	13a	2.56	2.59	1.2	2	10	2.82	2.74	3.1
N2	21	2.61	2.86	8.8	N2	13b	2.89	2.9	0.3	3	4a	2.34	2.35	0.5
N2	22	2.83	3.15	9.9	N2	21	3.38	3.55	4.8	3	4b	2.53	2.62	3.3
N3	N4	2.38	2.35	1	N2	22	2.83	2.87	1.3	3	10	3.05	3.07	0.6
N3	22	2.8	2.97	5.6	N3	N4	2.3	2.37	2.7	4b	10	2.86	2.9	1.4
N3	21	2.71	2.87	5.5	N3	22	2.94	2.96	0.6	6	7	2.16	2.32	6.9
N3	25	2.45	2.45	0	N3	24	3.36	3.45	2.6	12	15	2.92	2.8	4.2
N3	26/27	3.13	3.13	0	N3	25	2.88	2.99	3.6	21	22	2.65	2.78	4.6
N4	2	3.24	3.33	2.5	N3	26/27	3.78	3.79	0.2	24	25	2.57	2.66	3.4
N4	10	2.67	2.49	7.3	N4	2b	2.02	2.18	7.1	24	26/27	2.71	2.87	5.4
N4	24	2.96	2.97	0.4	N4	25	2.59	2.56	1.1	25	26/27	2.59	2.43	6.5
N4	25	3.07	2.68	14.7	N4	26/27	3.07	3.07	0					

(continued)

Table 3.21 (continued)

Azumamide A					Des-azumamide C					Epi-azumamide E				
H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%
N4	26/27	3.17	2.97	7	2b	2a	1.79[a]	1.79	0					
2	10	2.59	2.68	3.2	3	2a/4a/4b	1.99	2.07	3.8					
2	3	2.13	2.39	10.7	5	3	2.96	2.94	0.5					
2	4a	2.76	2.76	0	6	7	2.45	2.45	0.2					
4b	4a	1.82	1.82	0	12	15	3.02	2.86	5.7					
4a	10	2.71	2.97	8.7	13b	13a	2.08	1.7	22.1					
4b	10	3.52	3.51	0.3	13a	15	2.72	2.44	11.3					
12	15	2.65	2.65	0	13b	15	2.66	2.47	8					
13a	15	2.5	2.44	2.5	15	22	4.31	4.25	1.4					
13b	15	2.47	2.47	0	21	22	2.6	2.62	0.5					
21	22	2.77	2.77	0.1	24	25	2.44	2.56	4.8					
24	25	2.98	2.85	4.7	24	26/27	3.02	3.19	5.4					
24	26/27	2.62	2.81	6.9	25	26/27	2.74	2.59	5.7					
25	26/27	2.5	2.43	2.7										
MAE%				3.1	MAE%				3.5	MAE%				2.9
J-coupling constants					*J-coupling constants*					*J-coupling constants*				
H1	H2	Exp.	Calc.	Diff	H1	H2	Exp.	Calc.	Diff	H1	H2	Exp.	Calc.	Diff
N1	3	8	8	0	N1	3	7.3	7.4	0.1	N1	3	8.7	9.5	0.8
N2	12	9.1	8.9	0.2	N2	12	9.6	9.1	0.4	N2	12	8.9	10.9	2
N3	21	8.7	8.2	0.5	N3	21	9.7	9	0.7	N3	21	9.1	11.6	2.5
N4	24	8.3	8.3	0	N4	24	6.8	6.8	0	N4	24	8.5	9.4	0.9
2	3	3.8	3.8	0	2b	3	12.4	12.4	0	2	3	12.4	13	0.6

(continued)

Table 3.21 (continued)

Azumamide A					Des-azumamide C					Epi-azumamide E				
H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%	H1	H2	Exp.	Calc.	%
3	4a	8.1	8.1	0	2a	3	3.6	3.6	0	3	4a	4.6	4.6	0
3	4b	6.4	6.4	0	3	4a	6.5	6.5	0	3	4b	5.6	5.6	0
12	13a	9.4	8.8	0.6	3	4b	6.5	6.6	0.1	12	13a	10	10.5	0.5
12	13b	6.7	6.7	0	12	13a	9.3	9.3	0	12	13b	5.2	5.4	0.2
24	25	10.1	10.1	0	12	13b	5.8	5.8	0					
					24	25	6.7	6.8	0.1					

[a]Reference. Based on 2500 structures each, from MD simulations

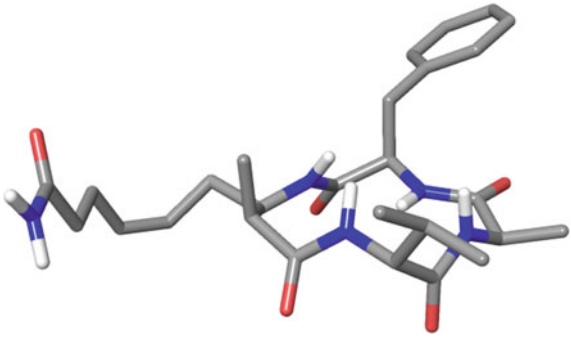


Fig. 3.21 Representational 3D structure from the constrained optimization of azumamide A. Structures of azumamide E are available from the literature for comparison [39, 40]

When investigating the NOE data, the constrained structure was in clear violation of multiple distances around the NH-2 site, which indicated that this structure was probably not that good a representation of the data, see Appendix A4.

Unconstrained structures

Unconstrained simulations were conducted in water and DMSO for the three azumamide analogues. The resulting conformational spaces were overlapping and the structures from water were used. Most distances and *J*-coupling constants were comparable between the three analogues, though some differences were observed. When fitting the NMR data to multiple structures this was translated to a mostly identical conformational space between the analogues. The correlations of the data to the simulated structures were good with MAE% of 3.1, 3.5 and 3.7% for Azu A, Des C and Epi E respectively. Also the *J*-coupling constants exhibited a good correlation between experimental and back-calculated values. The structures from unconstrained simulations were thus deemed to be better representations of the NMR data than the structures from constrained simulations and further conclusion were based on these. The experimental distances and back-calculated values, obtained by optimizing the correlation of 2500 structures from molecular dynamics iteratively, are shown in Table 3.21.

Structural differences

Using unconstrained simulations, the difference in the ring conformation between the three compounds was diminishing (Fig. 3.22). It would seem as if the natural compound had more flexibility in the ring, but the majority of the conformers which fitted the NMR data had NH-2 pointing upwards, in line with the synthetic structures. The best correlation to the data had NH-2 pointing upwards approximately 90% of the time, and downwards 10%. This is proposed to be translated to the intermediate position in the constrained calculations. A similar distribution was also seen for Des C. The multiple conformer approach led to a much better fit for the ring than the constrained simulations—still with errors around the NH-2 though, see Table 3.21.

The ring conformation was thus essentially unaffected by the inversion or removal of the methyl group, in good correlation to the NMR data and simulations.

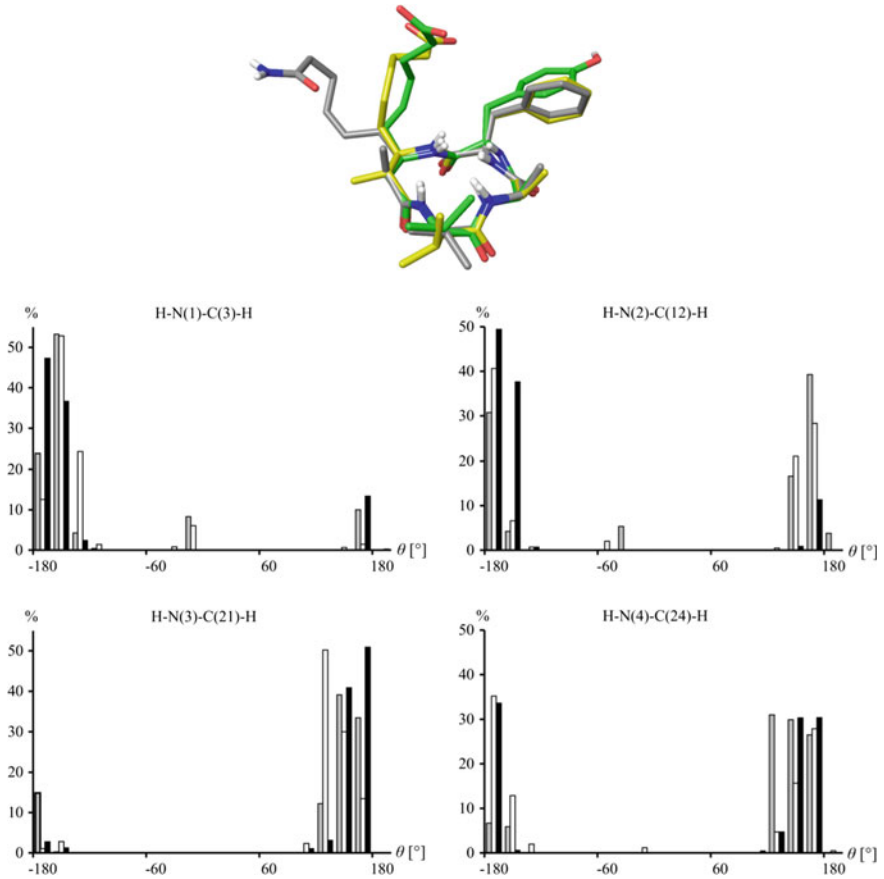


Fig. 3.22 Top: The best fit structure of the three analogues overlain; grey: Azu A, green: Des C, yellow: Epi E. Bottom: NH-C_{α}H dihedral angles of the conformers used in the NOE and *J*-coupling constant fit of azu A (grey), des C (white) and epi E (black)

This led to the search for, and identification of, other differences between the three analogues.

Side-chain orientation

The side-chain orientation was the only major difference between the different compounds. While the distances between the side-chains and the ring generally differed more than the ring proton distances, it was only translated into major differences in the orientation for the linker side-chain, which in turn is arguably the most important side-chain considering the involvement in reaching the binding site [37].

The side-chain of the linker was thus the only clear difference in NMR observables between the three compounds. This was also apparent from the simulations, but is more pronounced, and experimentally confirmed, in the NMR data.

The populations of the orientations for the different structural analogues are found in Fig. 3.23.

The natural compound, which is the most active, had an orientation of the linker side-chain centered on $\theta = \pm 60°$. The des-methyl analogue displayed an orientation shared between the three sites in a three jump model, while the epi-methyl, the least active analogue, displayed an orientation centered on $\theta = -80°$ and 180°. It may thus be proposed that favorable activity would be obtained when the orientation of the linker was at $\theta = +60°$ in the binding pocket, as this would correlate well to the activity data. This result was compared to results from the theoretical docking studies.

Docking results

The chosen HDAC to be used in the theoretical docking was HDA3, as a crystal structure of a co-repressor complex was available to dock into. The resulting azumamide structures had all amide protons directed to the same face and toward

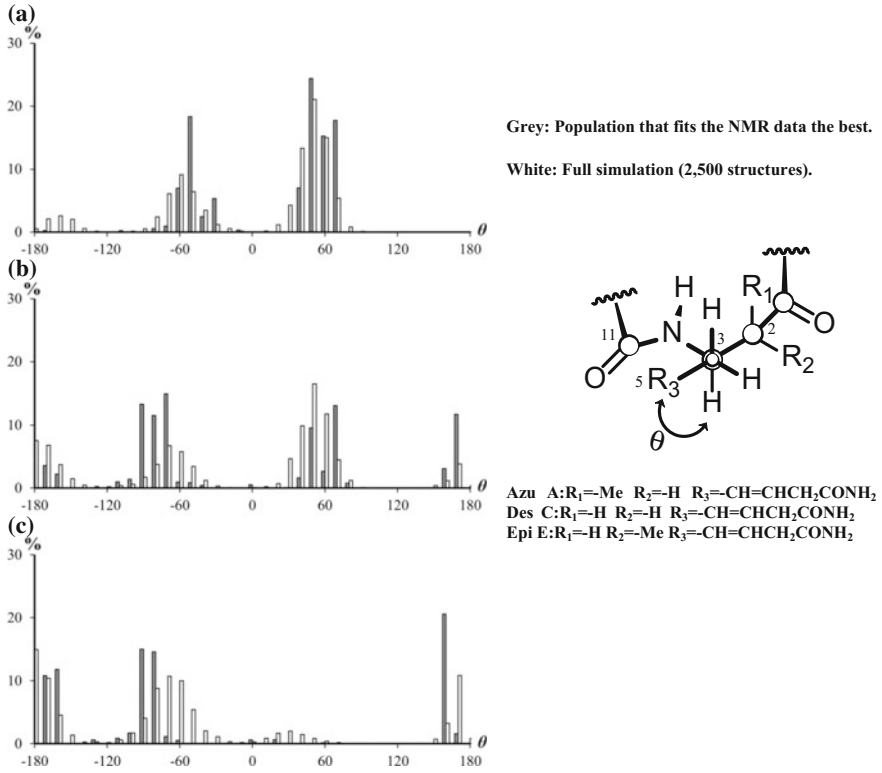


Fig. 3.23 The population in % of the orientations of the side-chain at C-3, illustrated by the dihedral angle H3-C3-C4-C5 (θ), for the natural azu A (**a**), and the synthetic des C (**b**) and epi E (**c**)

the carboxylate side-chain of Asp93, in agreement with solution data. The orientation of the aromatic side-chain also correlated well to the solution conformations. The linker side-chain had a dihedral angle of either 180° or 60°, except for azumamide E where a dihedral of 110° was prevalent. It was thus not possible to correlate the preferable dihedral angle from NMR to the docking studies, and since no docked crystal structure is available, a possible connection must be regarded as a possibility and supported speculations. It is of course not necessary for the bound compound to be of the same conformation as in the free liquid state, but it would lead to a potentially lower energy penalty upon binding. From theoretical calculations of the binding interactions, it was determined that the differences in activity should probably be contributed to a large methyl effect from interactions to Phe200 of the enzyme.

Minor conformer

It was apparent from the spectra that a minor conformational average was present in solution that averaged slowly on an NMR time scale from multiple resonances in the spectra. The data was thus catered for as discussed earlier for the natural compound CM-A. In contrast to CM-A, where the minor conformer was not investigated further, it was speculated that the minor conformer could be a *cis*-amide arrangement of one of the amide bonds in the ring. This is supported experimentally by an observed strong NOE between H-12 and H-21, leading to the *cis*-amide bond being between N-2 and C-20, the most flexible amide bond in the major conformational average. The 3D structures of these conformations were not determined due to weak NOE intensities. The differences between the ratios of the conformations were not significant between the different analogues and could thus not be used to explain the differences in the activity.

Conclusion

Solution state structures of natural and analogues of representative azumamides were investigated by NOE and *J*-coupling constants leading to knowledge of the structural space inhabited by the molecules. The ring structures of all were more or less equal with few differences between analogues. The orientation of the β^3-chain differed from natural azumamide to the unnatural structures, but this could not be correlated to theoretical docked structures. Still the knowledge of the structure was increased by applying a non-constrained methodology compared to previous publications, where a constrained simulation approach was utilized.

3.3.2 Molecular Recognition

Molecular recognition is defined as a specific non-covalent interaction of two or more molecules and is an essential part of many highly important biological functions, from protein folding, enzyme reactivity and our immune system to the structural organization of cells [42–44]. Due to the widespread dependence on molecular recognition in human biology, gaining further understanding of the

involved processes is important in the development of drugs that e.g. interfere with specific protein-protein interactions [45].

Ph.D. student Ming Li of the group of Professor Morten Meldal (MM), University of Copenhagen, developed a methodology of identifying peptide partners which bound with high specificity from on bead pairing of a large library of approximately 78,125 compounds. The bead-based screening protocol, developed in the lab of MM, allows for a simultaneous screening of up to 10^{12} peptide-peptide interactions on the biocompatible poly(acryloyl-bis-(aminopropyl)polyethylene glycol) (PEGA) resins [46, 47]. The screening was setup as to identify peptides which bound from electrostatic interactions, which was investigated using competitive binding condition, where salt concentrations were shown to influence the binding. Lastly the binding was evaluated by fluorescence microscopy using a custom-made flow-cell [48]. For more information regarding the comprehensive work of developing and utilizing the screening method, the reader is referred to the article [35]. Multiple binding partners were identified and a condensed conclusion of their findings is found in Table 3.22 which displays the best binding partners for the peptide Target 20.

The structural build of the best binding partners were all conserved, as amino acids with negatively charged side-chains in both ends and a single hydrophobic amino acid in the middle of the peptide, suggesting an electrostatic interaction. This is in contrast to many previous reports on molecular recognition, where hydrophobic interactions in water is usually the principal driving force towards binding [45, 49–52]. The *C*-terminal was kept a glycine due to easier coupling and cleavage of the peptide and resin.

NMR study

From the screening two peptides were chosen for structural elucidation; Target 20 (T20) and Ligand 7 (L7), shown in Fig. 3.24, which had shown a high degree of specific binding. T20 consists of ten amino acids while L7 consists of eight. The 3D structures of these compounds were investigated using distances from NOE data.

The two peptides display a specific binding interaction with a k_D of 0.7 μM. NMR data were obtained for the two molecules in isolation and of a mixture of the two. The assignments of the compound resonances are found in Appendix A4. Upon mixing the chemical shifts were almost unchanged, especially for T20, though some changes were observed in the form of differentiation of chemical shifts of overlapped peaks for L7, see Fig. 3.25. In contrast to the determined low binding constant, no intermolecular NOE correlations were observed. Some changes in distances of the individual peptides were observed upon mixing, but the changes were in most cases minor.

Table 3.22 Binding constants of various peptides (ligands) with Target 20

Ligand #	Structure	k_d (μM)
7	EDYEVEEG	0.658 ± 0.029
17	EDDWDDG	3.37 ± 0.25
18	EDYEVDDG	6.95 ± 1.49
19	EDYEWEEG	0.661 ± 0.059

Reproduced with permission from the publisher [35]

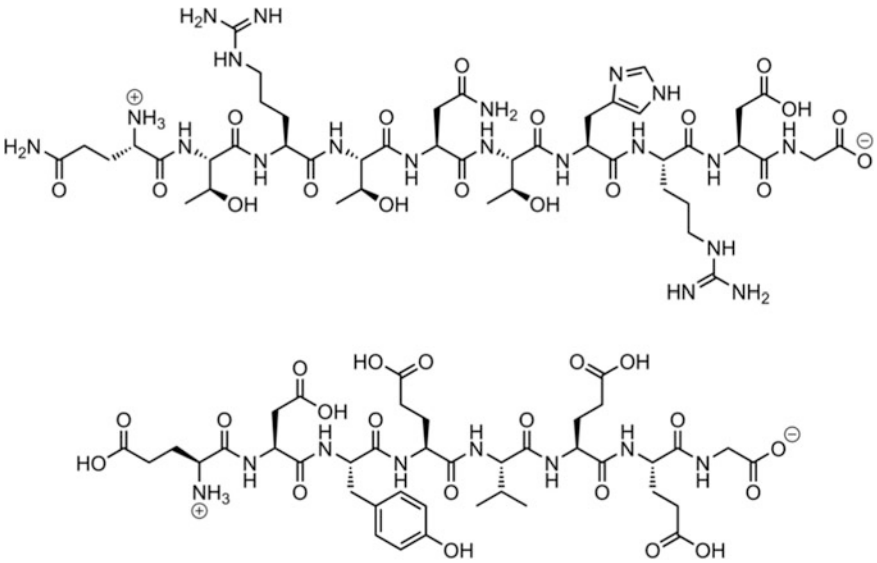


Fig. 3.24 Top: The structure of Target 20 (T20). Bottom: The structure of Ligand 7 (L7)

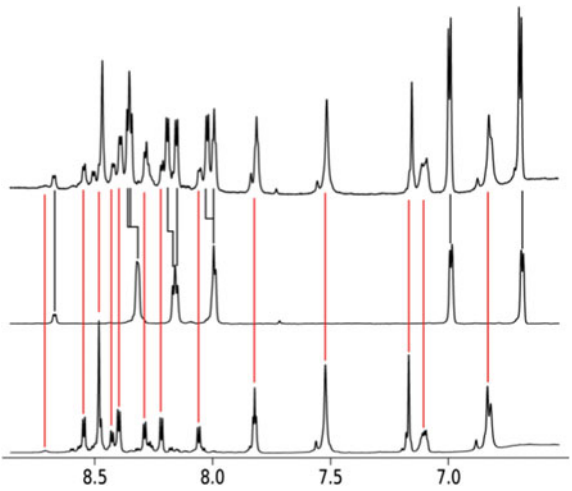


Fig. 3.25 Change in chemical shifts in the amide region upon mixing T20 and L7 at pH 6. The top spectrum is the mix, the middle spectrum is L7 and bottom spectrum is T20. Reproduced with permission from the publisher [35]

Differences in the rotational correlation time

An indication of interaction was an apparent increase in the rotational correlation time (τ_c) for both peptides upon mixing. This would be in correlation with an interaction so that both peptides behave as bigger molecules in solution. This may be shown utilizing the work of Macura et al. [Eqs. (2.3) and (2.11)], making (3.6) valid for short mixing times [53].

$$\frac{\eta_{IS}(t_m)}{1/2[\eta_I(t_m)+\eta_S(t_m)]} = -(W_{2IS} - W_{0IS})t_m = \sigma_{IS}t_m$$
$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2\gamma^2}{10}\left(\tau_c - \frac{6\tau_c}{1+4\omega^2\tau_c^2}\right) r_{IS}^{-6} \tag{3.6}$$

Equation (3.6) is a revisit from Sect. 2.1 and is valid for homonuclear NOEs, where ω is approximately $5 \cdot 109$ rad/s at 800 MHz. As seen, the value of σ_{IS} is dependent on the distance r and τ_c. For e.g. methylene proton pairs where the distance is known, the correlation time can be calculated. The shift in τ_c was immediately visible in the NOESY spectra as the cross-peak to diagonal-peak intensity was significantly higher for the mixture compared to the single peptides, while the summed concentration was equal and thus lower for the individual peptide. Since the NOEs were in-phase with the diagonal and thus higher than the crossover correlation time, an increase in the size of the peptide would increase the NOE intensities as given in Eq. (3.6) and thereby correlated to a longer τ_c (see Fig. 2.4, p. 7). It was observed that the σ for proton pairs generally increased, especially for T20, which translates to either a majority of the distances being shorter or a difference in τ_c; the latter being the most probable. Both scenarios indicated an interaction between the peptides though. For the methylene proton pairs τ_c could be calculated by assuming a fixed distance of 1.78 Å for methylene protons. Comparisons are found in Tables 3.23 and 3.24 for the Target, Ligand and the mix of the two.

It is of course possible that the rotational correlation time was increased by a difference in the viscosity of the solvent. It is thus important to state that the samples were acquired at the same temperature and pH and that any small

Table 3.23 Rotational correlation times (τ_c) and differences in cross-relaxation rate constants for protons in of T20, in complex with L7 and alone

H 1	H 2	σ_{IS} (Mix)	σ_{IS} (Single)	Increase (%)	τ_c (Single) (10^{-10} s^{-1})	τ_c (Mix) (10^{-10} s^{-1})
Arg(8)-βH2	Arg(8)-βH3	0.644	0.346	46	3.5	4.8
Thr(2)-NH	Gln(1)-αH	0.048	0.023	53		
Arg(3)-NH	Thr(2)-αH	0.107	0.212	−98		
Thr(4)-NH	Arg(3)-αH	0.100	0.101	−1		
Asn(5)-NH	Thr(4)-αH	0.150	0.063	58		
His(7)-NH	Thr(6)-αH	0.218	0.223	−2		
Arg(8)-NH	His(7)-αH	0.173	0.122	29		
Asp(9)-NH	Arg(8)-αH	0.229	0.125	45		
Gly(10)-NH	Asp(9)-αH	0.085	0.007	92		
His(7)-NH	Thr(6)-NH	0.115	0.133	−16		
Thr(6)-αH	Thr(6)-γH2	0.186	0.091	51		

Table 3.24 Rotational correlation times (τ_c) and differences in cross-relaxation rate constants for protons in of L7, in complex with T20 and alone

H 1	H 2	σ_{IS} (Mix)	σ_{IS} (Single)	Increase (%)	τ_c (Single) (10^{-10} s^{-1})	τ_c (Mix) (10^{-10} s^{-1})
Tyr(3)-βH2	Tyr(3)-βH3	0.783	0.490	37	4.1	5.4
Asp(2)-NH	Glu(1)-αH	0.126	0.096	24		
Asp(2)-NH	Asp(2)-αH	0.190	0.125	34		
Glu(6)-NH	Val(5)-αH	0.206	0.126	39		
Tyr(3)-NH	Asp(2)-αH	0.097	0.121	−25		
Val(5)-NH	Glu(4)-αH	0.245	0.082	67		
Gly(8)-NH	Glu(7)-αH	0.074	0.036	52		
Glu(4)-NH	Tyr(3)-αH	0.172	0.131	24		

differences in the two are not likely to give rise to the rather big increase in intensity. The possibility cannot completely be disregarded though.

Titration

To verify that a plausible interaction was present a titration study was performed, but no binding constant was obtainable, which would also be impossible to be studied with NMR due to a k_D far below a minimum of 10^{-5} suggested from the literature [54]. It was observed that the change in relative chemical shifts of the Ligand 7 was dependent on the presence of Target 20, by performing a titration of L7 with T20, as seen in Fig. 3.26, and afterwards diluting the sample to regain the original shifts.

Structures from MD simulations

The structures of Target 20 and Ligand 7 were identified to be too flexible to fit conformers from unconstrained simulations to the data due to too many and too different conformers. It was clear that some degree of order was present from multiple NH-C_αH distances corresponding or approaching those of β-strands (2.2 Å) or long distances for both compounds, given in Table 3.25, and constrained

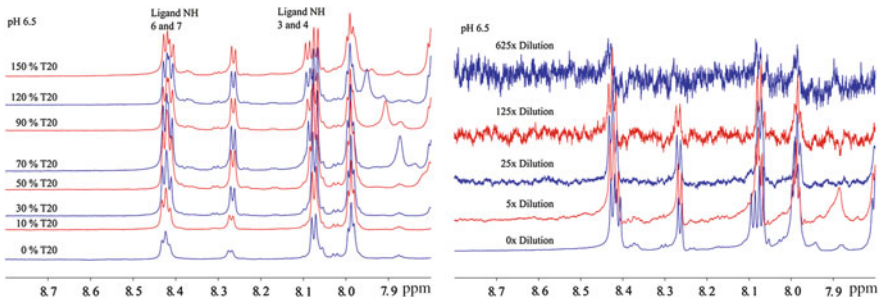


Fig. 3.26 Left: 1D ^{1}H spectra of the titration of L7 with T20 at pH 6.5. Note that the resonances of T20 were not observed at this pH due to exchange with the solvent. Right: dilutions of the sample after titration

Table 3.25 Experimental distances of NH_i-$C_\alpha H_{i-1}$ for T20 and L7, as individual peptides and when mixed

H1	H2	Dist. (Å)	H1	H2	Dist. (Å)
Target 20, individual peptide			*Target 20, mixture*		
Thr(2)-NH	Gln(1)-αH	3.21	Thr(2)-NH	Gln(1)-αH	2.94
Arg(3)-NH	Thr(2)-αH	2.42	Arg(3)-NH	Thr(2)-αH	2.50
Arg(3)-NH	Thr(4)-NH	3.29	–	–	–
Thr(4)-NH	Arg(3)-αH	2.38	Thr(4)-NH	Arg(3)-αH	2.54
Asn(5)-NH	Thr(4)-αH	2.43	Asn(5)-NH	Thr(4)-αH	2.37
His(7)-NH	Thr(6)-αH	2.30	His(7)-NH	Thr(6)-αH	2.22
His(7)-NH	Thr(6)-NH	2.55	His(7)-NH	Thr(6)-NH	2.38
Arg(8)-NH	His(7)-αH	2.16	Arg(8)-NH	His(7)-αH	2.31
Asp(9)-NH	Arg(8)-αH	2.46	Asp(9)-NH	Arg(8)-αH	2.21
Gly(10)-NH	Asp(9)-αH	3.19	Gly(10)-NH	Asp(9)-αH	2.60
Ligand 7, individual peptide			*Ligand 7, mixture*		
Asp(2)-NH	Glu(1)-αH	2.44	Asp(2)-NH	Glu(1)-αH	2.44
Tyr(3)-NH	Asp(2)-αH	2.35	Tyr(3)-NH	Asp(2)-αH	2.55
Glu(4)-NH	Tyr(3)-αH	2.32	Glu(4)-NH	Tyr(3)-αH	2.31
Glu(4)-NH	Val(5)-NH	2.96	Glu(4)-NH	Val(5)-NH	2.98
Val(5)-NH	Glu(4)-αH	2.51	Val(5)-NH	Glu(4)-αH	2.18
Glu(6)-NH	Val(5)-αH	2.33	Glu(6)-NH	Val(5)-αH	2.25
Glu(6)-NH	Val(5)-NH	3.04	Glu(6)-NH	Val(5)-NH	2.89
Gly(8)-NH	Glu(7)-αH	2.88	Gly(8)-NH	Glu(7)-αH	2.66
Gly(8)-NH	Glu(7)-NH	2.99	–	–	–

Determined by setting the distance of a diastereotopic proton pair to 1.78 Å (T20: Arg(8)-β, L7: Tyr(3)-β)

simulations were used in order to get backbone structures which correlated well to the NOE data [55].

The resulting backbone structures fit well to the data and are shown in Fig. 3.27 but it is imperative to stress that these are representations of the conformational space.

The two structures both exhibited a turn, in good agreement with NH-NH distances from the NOE data. The two structures were found to undergo small adaptations upon interacting, but the changes were minor, and largest for L7, again in good correlation to the NMR data, illustrated in Fig. 3.28. Four water molecules were entrapped in a cavity between the two structures in the simulations, which established a hydrogen bonding network with both T20 and L7, while electrostatic interactions and topological complementarity between the structures were observed along the rim of the cavity. Note that the water molecules are not shown in Fig. 3.28. The interaction was thus based on side-chain to side-chain interactions, and water mediated hydrogen bonds, which could be the reason for the lack of intermolecular NOEs. Another suggestion could be a high on-off rate which could hamper the buildup of NOE signal.

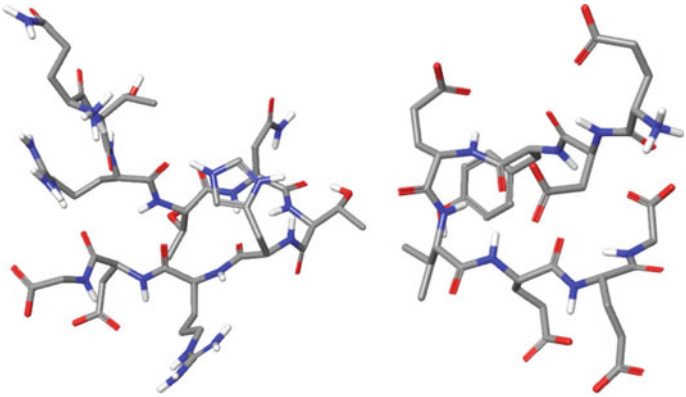


Fig. 3.27 Representative 3D structure of T20 (left) and L7 (right) from constrained simulations using NOE-derived distances

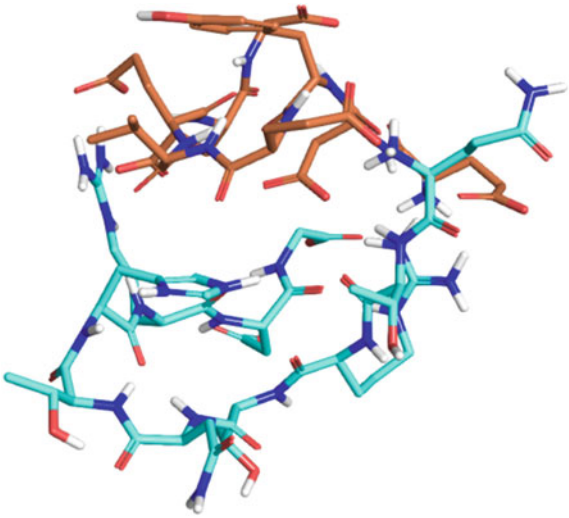


Fig. 3.28 The interaction between target 20 (teal) and ligand 7 (orange) from MD simulations using constraints from NOE distances. Interactions are dominated by charged side-chain interactions. Water molecules in the cavity of the peptides omitted

Conclusion

Two peptides were investigated by NMR spectroscopy to obtain information of a highly specific binding interaction, suggested from binding data. It was not possible to identify the interaction from inter-peptide NOEs, but several indices suggested that the peptides interact, including shift in proton chemical shifts upon mixing, a shift in the rotational correlation time and small changes in NOE intensities. The peptides were modelled based on distances obtained from the relative NOE intensities, and two complementing structures were identified, which fitted the backbone data well, and also theoretically bind well, primarily by interactions of the outermost parts of side-chains. The fact that the peptides contained a potential for

much more flexibility, led to the need for constrained simulations to obtain good structures, and the given structures are given as representatives of the conformational space the peptides occupy.

3.4 Other Compounds

This and the following Section cover a couple of projects and points from my external stay in the group of Craig Butts at the University of Bristol. This first section covers the usage of NOEs coupled to DFT optimized or clustered structures to gain knowledge of conformer populations. The NOE work is based on the PANIC approach, covered in Sect. 2.1.5.

3.4.1 Population Analysis of Quinine

The alkaloid quinine has been thoroughly investigated due to interesting biological activities and usage in chiral synthesis [56–59]. The conformational populations have previously been solved from Gibbs free energies from Hatree-Fock and DFT calculation and the *J*-coupling constants of H8–H9 [60, 61]. The related compound cinchonidine, varied only by missing the methoxy group at the aromatic position 6' (Fig. 3.29), was recently investigated using DFT calculations and NOEs, although a limited set of NOEs was used (only <3 Å included) and the conclusions were negative in correlating the theoretical energy populations to the NOE populations [62]. It could thus be interesting to see if a conclusion could be reached for quinine by the usage of the PANIC approach and a full NOE dataset (Table 3.26). When investigating the spectra only three distances above 3 Å were available, all of which were fixed distances due to structural rigidity, so the difference between datasets is not expected to be significant.

The group of Craig Butts had previously investigated quinine and determined the conformers from relative conformer energies at a B3LYP/6-31 g(d) level of theory in implicit chloroform, and the conformers are compared to those reported by Urakawa et al. for cinchonidine, see Table 3.27 and Fig. 3.30 [62].

Most of the distances were between rigid parts of the molecule, and should not influence the determined populations. The non-rigid distances of the seven relevant

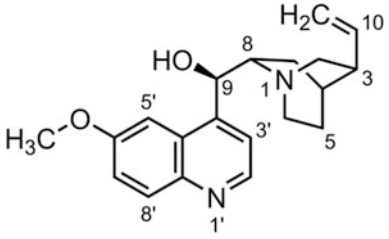


Fig. 3.29 Structure of quinine

Table 3.26 Distances used in the conformational analysis of quinine

H1	H2	Exp. Dist. (Å)
2′	3′	2.42
9	3′	2.85
9	5′	2.11
9	6a	2.51
9	8	2.55
9	7a	2.82
6a	6b	1.79[a]
6a	5a	2.31
6a	5b	3.10
8′	7′	2.54
3	2a	2.36
3	6b	3.14
3	4	2.55
3	5b	2.49
8	3′	2.76
8	5′	2.36
2a	2b	1.71
8	7a	2.82
8	7b	2.29
2b	3	2.92
6b	5a	3.11
6b	5b	2.38

[a]Reference

conformers which influenced the obtained populations are given in Table 3.28. All distances are included in Appendix A4.

Initially the populations of the six conformations identified in the group of Craig Butts were compared to the NOE data. The populations of single conformers were fixed individually and the remaining conformers were optimized to give the best fit to the data. It was realized that the MAE was not discriminative, since the MAEs of most populations were generally low. Instead the maximum difference for the populations was used, as a better discriminator of the actual correlation between NOE distances and 3D structures see Fig. 3.31.

NOE distances dictated that the O3 conformer was indeed the dominating one, in accordance with all DFT calculations, and that the remaining conformers were populated 0–20%. Due to overlap in the conformational space between the conformers only three conformers were needed in order to obtain a good fit to the data, to a good approximation, namely O3, C1 and C7. There was no solid experimental evidence that the rest of the conformations are present in solution, probably due to structural similarities within open and closed conformations.

Table 3.27 Populations in % from the literature and from the group of Craig Butts

	Dihedral angles (°)			% Population	
Conformer	τ_1	τ_2	τ_3	Urakawa [62]	Butts (Unpublished work of the group of Dr. Craig Butts)
C1	249	50	175	10	8
C2	60	60	180	4	2
C7	17	53	173	11	10
O3	99	150	278	72	73
O4	260	140	272	0	4
O8	20	150	278	0	3
O10	102	268	38	3	0

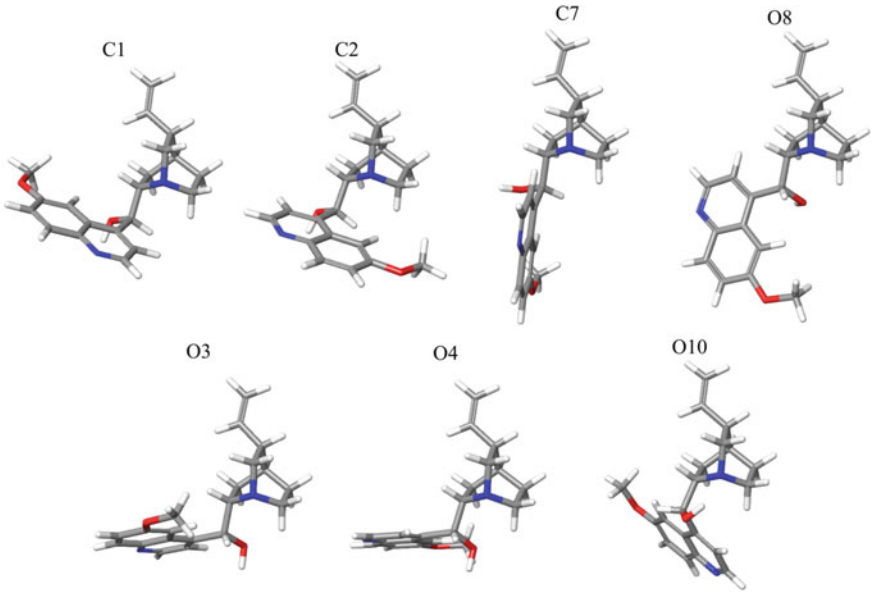


Fig. 3.30 The most abundant conformers of quinine (B3LYP/6-31 g(d) and literature [62])

When adding the seventh conformation, O10, to the available structures as given by the population analysis of Urakawa et al., it led to a slightly better fit, considering the MAE, while the maximum difference in the distances was almost unaffected. The NOE populations were generally worse compared to the theoretical populations, and akin to C2, O4 and O8, the conformer was not proven needed in the fitting due to a low dependence on the inclusion of the conformer, Fig. 3.32.

In Table 3.29 the conformational populations using the different sets of structures are shown, which may be compared to the data in Table 3.27.

Table 3.28 Experimental distances between non-rigid proton pairs and distances determined from the seven relevant conformers of quinine

H1	H2	Exp.	C1	C2	O3	C7	O8	O4	O10
9	3′	2.9	2.28	3.68	3.65	3.55	3.62	3.62	3.61
9	5′	2.1	3.79	2.09	2.13	2.34	2.29	2.29	2.07
9	6a	2.5	2.13	2.16	2.67	2.23	2.69	2.69	3.97
9	8	2.6	3.04	3.04	2.56	3.05	2.53	2.53	2.24
9	7a	2.8	2.60	2.71	3.63	2.44	3.65	3.65	2.85
8	3′	2.7	2.75	2.72	3.83	2.33	2.19	2.19	2.45
8	5′	2.4	2.41	2.22	2.44	3.98	3.70	3.70	3.30

Structures were optimized using MPW1PW91/6-311 + g(d,p)

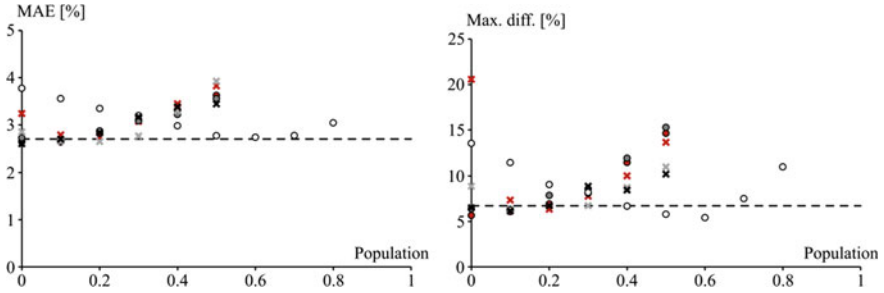


Fig. 3.31 Population analyses, by fixing a conformer population and optimizing the rest of the six conformers C1(×), C2(×), C7(×), O3(○), O4(●) and O8(●). The dotted line is the value using only C1, C7 and O3

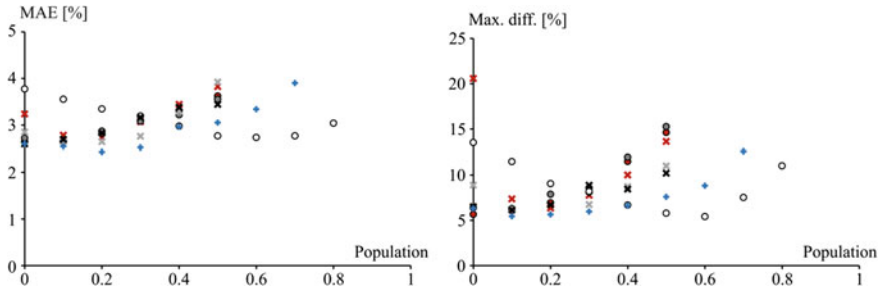


Fig. 3.32 Population analyses, by fixing a conformer population and optimizing the rest of the seven conformers C1(×), C2(×), C7(×), O3(○), O4(●), O8(●) and O10(+)

In conclusion it was difficult to determine the populations from the data, as the introduction of the O10 conformer led to large differences between experimental and theoretical populations. The problems arise from overlaps of the conformational space between conformers, leading to few distances that may actually

Table 3.29 Experimental population of the conformers from NOE distances, using different sets of conformers as indicated

	Exp. population (%)		
Conformer	6 structures	7 structures	Essential 3
C1	15	16	13
C2	7	16	–
C7	14	4	21
O3	59	41	66
O4	4	0	–
O8	1	0	–
O10	–	23	–
MAE (%)	2.54	2.43	2.70
Max diff (%)	6.31	5.45	6.73

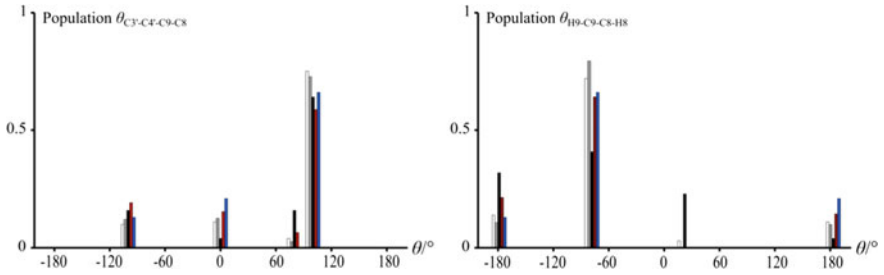


Fig. 3.33 The population of dihedral angles associated with the rotatable bonds in quinine. White: Urakawa (energy), grey: Butts (energy), black: 7 structures (NOE), red: 6 structures (NOE), blue: 3 structures (NOE)

differentiate conformers. There were seven distances identified between rotatable parts of the structure, which, compared to up to seven conformers, lead to an underdetermined system when trying to determine a population of each conformer. It might be more interesting to fit the dihedral angles around the rotatable bonds, which is seen in Fig. 3.33.

The only really different populations were observed when including the seventh conformer which seemed to be severely overestimated. The populations of quinine are revisited in Chap. 8.

3.5 Inclusion of Correlation Time in NOE Calculations

On my external stay I was involved in a project where the goal was to utilize calculation of correlation times to improve on the accuracy of NOE calculations for small organic molecules. This was tried achieved in multiple ways, I will include only the parts where I was directly involved.

3.5.1 Theory

As a major assumption in the ISPA methodology, is equality of correlation time (τ_c) for all nuclei pairs, it would seem that including relative correlation times would increase the accuracy of the calculations, as proposed in the literature [63–68].

The actual parameter of interest is the effective rotational correlation time, which covers the overall rotational correlation time of the molecule (τ_o) and the relative intramolecular movements (τ_s) as given in (3.7) [63].

$$\frac{1}{\tau_c} = \frac{1}{\tau_o} + \frac{1}{\tau_s} \tag{3.7}$$

Note that the effective correlation time is also called τ_e or τ_i in the literature but covers the same parameter [63, 64].

Values of τ_c may be obtained in four different ways: [69, 70]

1. By estimation from the size of the molecule (correlates with molecular weight of proteins).
2. By estimating correlation times from T_1 relaxation (only possible for ^{13}C).
3. Utilizing NOE and ROE spectra at the same field strength.
4. Comparing NOEs at two field strengths.

The first two methods will not be discussed while the formula needed for the two latter methods are included in Eqs. (3.8) to (3.16), starting from the differences in NOE and ROE observables in (3.8) and (3.9) [63–65, 69]. Note a small overlap with theory in Chap. 2.

$$\sigma_{NOE} = K \cdot r^{-6} \cdot [6J(2\omega) - J(0)] = k \cdot r^{-6} \cdot \tau_c \left(\frac{6}{1 + 4\omega^2 \tau_c^2} - 1 \right) \tag{3.8}$$

$$\sigma_{ROE} = K \cdot r^{-6} \cdot [3J(\omega) + 2J(0)] = k \cdot r^{-6} \cdot \tau_c \left(\frac{3}{1 + \omega^2 \tau_c^2} + 2 \right)$$
$$K = \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\hbar^2 \gamma_H^4}{10} \tag{3.9}$$

In practice the most utilized ROE experiment is the T-ROE, due to the less demanding spin lock utilized. In T-ROE the magnetization may be viewed as shared between longitudinal and transverse during the mixing time as given in (3.10) [71, 72].

$$\sigma_{T-ROE} = \frac{\cos\theta_i \cos\theta_j \sigma_{NOE} + \left(1 + \sin\theta_i \sin\theta_j\right) \sigma_{ROE}}{2} \tag{3.10}$$

With $\tan\theta = (\omega_i - \omega_T)/\gamma B_1$. ω_i is the chemical shift of i in rad/s (or Hz, need to match γB_1), ω_T is the chemical shift of spin-lock transmitter offset in rad/s (or Hz), γB_1 is the spin-lock field in rad/s (or Hz). If the assumption is made that θ_i and

$\theta_j \rightarrow 0$, then $\cos\theta_i$ and $\cos\theta_j \rightarrow 1$ and $\sin\theta_i$ and $\sin\theta_j \rightarrow 0$. This is the most abundant approach used in the literature on the subject [63–68], and leads to (3.11).

$$\begin{aligned}\sigma_{T-ROE} &= \frac{K}{2} \cdot r^{-6} \cdot [6J(2\omega) + 3J(\omega) + J(0)] \\ &= \frac{K}{2} \cdot r^{-6} \cdot \tau_c \left(\frac{6}{1+4\omega^2\tau_c^2} + \frac{3}{1+\omega^2\tau_c^2} + 1 \right)\end{aligned} \tag{3.11}$$

It is fairly obvious that this assumption is not valid for most systems. If one only wants to utilize the T-ROE intensities for distance calculations the approximation may work just fine due to the aforementioned r^{-6} relationship, but the influence on τ_c may be problematic. With a γB_1 of 5500 Hz at 500 MHz, a proton 4 ppm from the offset with its "partner" proton on the offset resonance will need a correction of σ_{ROE} of $\approx 5\%$. If the two protons reside at ± 4 ppm of the offset resonance a correction of σ_{ROE} of $\approx 25\%$ is needed.

From Eqs. (3.8) to (3.11) the ratios of NOE to (T-)ROE may be given as Eqs. (3.12) and (3.13) [64, 73].

$$\frac{\sigma_{NOE}}{\sigma_{ROE}} = \frac{5 + \omega^2\tau_c^2 - 4\omega^4\tau_c^4}{5 + 22\omega^2\tau_c^2 + 8\omega^4\tau_c^4} \tag{3.12}$$

$$\frac{\sigma_{NOE}}{\sigma_{T-ROE}} = \frac{10 + 2\omega^2\tau_c^2 - 8\omega^4\tau_c^4}{10 + 23\omega^2\tau_c^2 + 4\omega^4\tau_c^4} \tag{3.13}$$

Note again that the latter is only true if θ_i and $\theta_j \rightarrow 0$. Still, even with this restriction, this is what is utilized in the literature [63, 64, 73]. Experimental data that suggests the need for the correction is presented in Appendix A9 in the form of different, albeit correctable, obtained T-ROE intensities at different pulsed field strengths and offsets. The relative NOE to ROE intensities are illustrated in Fig. 3.34.

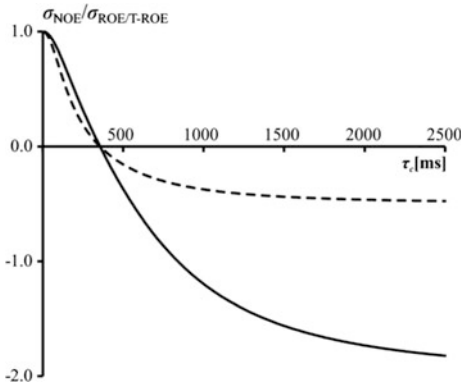


Fig. 3.34 The theoretical relative cross-relaxation rate (σ) of NOE and ROE (—) or T-ROE (- -) experiments at 500 MHz for a proton pair. Calculated using (3.12) and (3.13)

If NOE and T-ROE data is acquired the corrected ROE part of T-ROE can be extracted by (3.14)

$$\sigma_{ROE} = \frac{2\sigma_{T-ROE} - \cos\theta_i \cos\theta_j \sigma_{NOE}}{\left(1 + \sin\theta_i \sin\theta_j\right)} \tag{3.14}$$

From (3.12) and (3.13) the correlation time may be calculated by (3.14) and (3.15) for $\sigma_{NOE} > 0$ and defining that $\sigma_{NOE}/\sigma_{(T-)ROE} = \nabla\sigma$.

$$\tau_c^{NOE/ROE} = \frac{1}{4}\frac{\sqrt{2}\sqrt{(2\nabla\sigma+1)(-22\nabla\sigma+1+3\sqrt{36\nabla\sigma^2+4\nabla\sigma+9}}}{(2\nabla\sigma+1)\omega} \tag{3.15}$$

$$\tau_c^{NOE/T-ROE} = \frac{1}{4}\frac{\sqrt{2}\sqrt{(\nabla\sigma+2)(-23\nabla\sigma+2+3\sqrt{41\nabla\sigma^2-28\nabla\sigma+36}}}{(\nabla\sigma+2)\omega} \tag{3.16}$$

For two field NOE analysis only (3.8) is needed. If data is acquired at field x and y, and assuming $\sigma_{NOE,x} > \sigma_{NOE,y}$, Eqs. (3.17) to (3.19) may be used [68].

$$\sigma_{NOE,x} = K \cdot r^{-6} \cdot \tau_c \left(\frac{6}{1+4\omega_x^2\tau_c^2} - 1\right) \tag{3.17}$$

$$\sigma_{NOE,y} = K \cdot r^{-6} \cdot \tau_c \left(\frac{6}{1+4\omega_y^2\tau_c^2} - 1\right) \tag{3.18}$$

$$\frac{\sigma_{NOE,x}}{\sigma_{NOE,y}} = \frac{\left(4\omega_x^2\tau_c^2 - 5\right)\left(4\omega_y^2\tau_c^2 + 1\right)}{\left(4\omega_x^2\tau_c^2 + 1\right)\left(4\omega_y^2\tau_c^2 - 5\right)} \tag{3.19}$$

From this the correlation time may be determined.

After the correlation time is determined the distances need to be calculated. There are different possibilities in doing this; first by using (3.8) to (3.11) directly, as all parameters but the distance is known. Another method is using a PANIC like approach, exemplified in (3.20).

$$r_i = r_{ref}\left(\frac{\sigma_i}{\sigma_{ref}}\frac{\tau_{c,ref}\left(\frac{6}{1+4\omega^2\tau_{c,ref}^2} - 1\right)}{\tau_{c,i}\left(\frac{6}{1+4\omega^2\tau_{c,i}^2} - 1\right)}\right)^{-1/6} \tag{3.20}$$

It should be noted that in most cases where correlation times are used, they are not included in the distance calculations but compared and deemed identical. It is thus used more to validate that the ISPA approach is appropriate [64, 73].

For the sake of inclusion, the last major equation which should be mentioned in regard to rotational correlation time is based in the rotational correlation times for ^{13}C which may be calculated from the T_1 relaxation time as shown in (3.21) [69].

$$\frac{1}{T_1(^{13}C)} = \left(\frac{\mu_0}{4\pi}\right)^2 N\hbar^2\gamma_H^2\gamma_C^2 r^{-6}\tau_c \tag{3.21}$$

where r is the C-H distance ($\sim$1.1 Å) and N is the number of attached protons. The usage of this formula will of course approximate that the, at least relative, rotational correlation times are equal for bound protons and carbons, which may be a good approximation, though it was not thoroughly investigated.

3.5.2 Inclusion of Correlation Times for Small Molecular NMR

To test if the inclusion of the rotational correlation time would improve the correlation between experimental NOE derived distances and distances from 3D structures of small organic molecules in organic solvents, three compounds were used; rigid, and more importantly here, spherical strychnine, rigid but flat progesterone and more flexible quinine. While data of strychnine was only acquired in chloroform, data for the remaining two compounds were also acquired in DMSO.

Strychnine

Due to the spherical structure and excellent fit between experimental and 3D structure distances of strychnine it was not anticipated that including the rotational correlation time would lead to an overall better structural distance correlation. Experimentally, many of the observed T-ROEs were of lower intensity than the NOE counterparts, which is not theoretically possible, if assuming that the formulae in the previous section are valid. This is problematic and results in a significantly smaller number of NOEs which are compared in Table 3.30. The fit between experimental and 3D structure distances was not improved but seems to be corrected more or less at random.

It was realized that the problem was the low values of the rotational correlation time for strychnine. Thus the theoretical differences between NOE and T-ROE intensities were lower than the experimental error. This is in good agreement with the observations; the values of τ_c were probably in the correct interval, but small differences were not translated well to distances. This problem is further addressed and explained in Table 3.31. Remember that a resulting distance error of 2 or 5% translates to the intensity theoretically being either $\sim$11 or 25% low or high, or $\sim$13 or 36% too high. The values or 2 and 5% were chosen as the average error approaches 2% for strychnine and the maximum single distance error approaches 5%. The errors in observed intensities were thus easily above the differences from differences in rotational correlation times. For comparison an intensity error of 2%

Table 3.30 Theoretical and experimental distances between protons in strychnine using NOEs or τ_c corrected distances using the NOE/T-ROE methodology (τ_m = 300 ms)

H1	H2	3D struct. (Å)	Experimental distances (Å)				τ_c (ps)
			NOE	Diff (%)	NOE/T-ROE	Diff (%)	
4	3	2.49	2.55	2.26	2.52	1.10	31.1
13	8	3.00	3.17	5.74	3.27	9.09	41.5
13	12	2.35	2.34	0.24	2.45	4.27	45.6
13	14	2.41	2.38	1.32	2.55	5.68	55.1
13	15a	2.24	2.31	2.73	2.28	1.65	31.3
15a	13	2.24	2.23	0.71	2.29	1.83	39.8
15a	15b	1.75[a]	1.75	0.20	1.75	0.20	33.5
15a	14	2.54	2.52	0.89	2.55	0.09	35.8
15a	16	2.47	2.44	0.98	2.48	0.66	37.5
18b	8	2.30	2.19	4.79	2.22	3.43	36.9
18b	18a	1.77	1.82	3.01	1.79	1.34	30.1
18b	22	3.12	2.99	4.30	3.20	2.35	54.5
			MAE (%)	2.26		2.64	

Structure optimized to MPW1PW91/6-311 + g(d,p). Only distances where NOE < T-ROE intensities are included. [a]Reference distance

Table 3.31 Theoretical PANIC intensities of different distances at 500 MHz assuming τ_c = 30 ps

r	η_{NOE}	$\eta_{T\text{-}ROE}$	Diff (%)
1.75	128.8	131.2	1.9
2.25	30.2	30.8	2.0
2.50	13.6	13.9	2.2

translates to a difference in 0.33% in distances, which is the precision needed for NOE calculations to obtain a better fit in the case of strychnine.

Slightly discouraged by this result, the fact that strychnine is spherical and that chloroform has a low viscosity and thus favors short τ_c's, prompted for the investigation into other compounds which could be investigated in multiple solvents, chosen to be chloroform and DMSO due to large differences in viscosity. Salts of strychnine may be dissolved in DMSO, which was attempted, but low solubility gave inconclusive results that are not included.

Progesterone

The second compound to be investigated was the steroid progesterone. The structure of progesterone is flat and elongated in virtue of being a steroid. It was thus proposed that the rotational correlation time would differ significantly dependent on whether the distance vector would lie along the long (L or 1), short (S or 2) axis or in plane (3), see Fig. 3.35 for clarification.

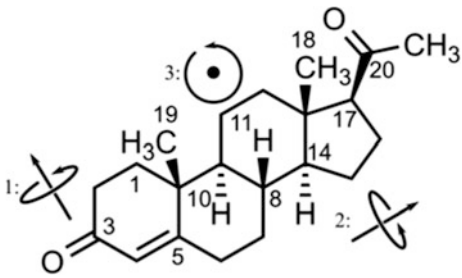


Fig. 3.35 The structure of progesterone, with the three proposed major rotation axis indicated. In the following the term "L" will cover 1 and 3, and "S" 2

A theoretical dependence was identified as two distinct groups, correlating to rotation 1 and 2, the first also including 3, with grouped distance errors. In practice it was troublesome to couple this to the actual rotational correlation time in chloroform, and only minor improvements were obtained when tested in DMSO.

Chloroform

The data obtained from chloroform is seen in Table 3.32. The direction of the distance vector were divided into long (L) and short (S) from assessment of a 3D structure (see Fig. 3.35, where L = 1 and 3, and S = 2). The theoretically optimized experimental distances (Exp Theo) give experimental distances from theoretically optimized values of τ_c of these two groups which identified to be L = 37 and S = 20 ps (from optimizations using the Excel solver function), which give a correction factor of 1.10 between the two. The experimental distances were

Table 3.32 Progesterone in $CDCl_3$

H1	H2	Dir	Calc.	Exp.	Diff	Exp. Theo.	τ_c (ps)	Exp. Corr.	Impr.
4	6a	L	2.36	2.15	0.21	2.36	36	2.22	0.07
17	16a	S	2.27	2.32	−0.05	2.32	21	2.21	−0.02
17	12a	L	2.43	2.24	0.19	2.46	42	2.37	0.13
17	14	L	2.46	2.16	0.29	2.38	40	2.27	0.10
6a	4	L	2.36	2.04	0.32	2.24	38	2.12	0.08
6a	7b	L	2.46	2.16	0.30	2.37	54	2.35	0.19
6a	7a	L	2.45	2.32	0.12	2.55	46	2.48	0.09
16b	16a	S	1.76	1.75	0.01	1.75	60	1.93	−0.17
16b	15b	L	2.31	2.19	0.12	2.40	97	2.52	−0.10
7b	6b	L	2.44	2.33	0.11	2.56	45	2.48	0.06
7b	6a	L	2.46	2.20	0.26	2.41	63	2.44	0.24
7b	8	S	2.46	2.48	−0.02	2.48	76	2.81	−0.33
7b	15b	L	2.78	2.47	0.31	2.71	64	2.75	0.28
7b	7a	S	1.76	1.74	0.03	1.74	64	1.93	−0.15

(continued)

Table 3.32 (continued)

8	6b	L	2.58	2.34	0.24	2.57	37	2.43	0.09
8	7b	S	2.46	2.56	−0.10	2.56	86	2.93	−0.37
8	15b	L	2.68	2.56	0.12	2.81	-*		
15b	16b	L	2.31	2.16	0.15	2.37	76	2.45	0.01
15b	7b	L	2.78	2.34	0.44	2.57	-*		
15b	15a	S	1.77	1.77	0.00	1.77	-*		
15b	8	L	2.68	2.48	0.20	2.72	61	2.75	0.13
7a	6a	S	2.45	2.43	0.02	2.43	27	2.41	−0.03
7a	7b	S	1.76	1.76	0.00	1.76	29	1.76	0.00
7a	15a	L	3.05	2.64	0.41	2.90	92	3.03	0.39
9	1b	L	2.29	2.08	0.20	2.29	42	2.20	0.11
MAE (%)				6.63		2.52		5.50	
St. dev. (%)				4.71		2.03		4.76	

NOESY and T-ROESY data used for calculation of τ_c (τ_m = 300 ms). Structure optimized to MPW1PW91/6-311 + g(d,p). *NOE/ROE was not solvable as $I_{NOE} > I_{ROE}$. H7b-H7a was used as reference. See text for further explanations

corrected by the τ_c determined from the NOE/ROE intensities (Exp. Corr.). The differences between experimental and theoretical distances for corrected compared to uncorrected data are given by subtracting the error of the corrected data from the uncorrected (Impr.); positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data.

The directionality of the distance vectors were assessed from a model and some may correspond more to a situation represented by rotation 3 in Fig. 3.35. Still, when comparing the experimental to the theoretical distances, there seems to be a correlation between the assigned rotation axis and the fit. This was translated into theoretical τ_c values, and led to a much improved fit of the experimental distances. In practice the determined τ_c seems to be distributed at random, and though the experimental distances generally correlate better to the 3D structure, multiple distances had increased differences to the 3D structure. Also, as for strychnine, some τ_c values were not possible to identify due to a lower ROESY intensity compared to the relevant NOESY signal, which results in unsolvable equations. This may, as for strychnine, be caused by the experimental error being larger than the theoretical difference between the NOESY and ROESY intensities, see Sect. 3.5 for further discussion.

DMSO

The data obtained from DMSO is found in Table 3.33.

As for chloroform the directionality of the distance vectors were assessed from a model. In DMSO it is less obvious that one may correlate experimental distances to markedly different τ_c. It was done though and led to an improvement in the already excellent fit between experimental distances and distances obtained from a 3D structure. In practice the determined τ_c seems to be distributed at random again, and

Table 3.33 Progesterone in DMSO-d_6

H1	H2	Dir	Calc.	Exp.	Diff	Exp. Theo.	τ_c (ps)	Exp. Corr.	Impr.
4	6a	L	2.29	2.25	0.04	2.31	85	2.22	−0.02
17	12b	S	2.73	2.84	−0.11	2.84	83	2.81	0.03
17	16a	S	2.36	2.36	0.00	2.36	95	2.36	0.00
17	12a	L	2.37	2.34	0.03	2.40	105	2.34	0.01
17	14	L	2.36	2.27	0.09	2.34	138	2.28	0.01
2b	1a	S	2.49	2.49	0.00	2.49	135	2.50	−0.01
6b	7b	L	2.44	2.39	0.06	2.45	96	2.38	0.00
6b	8	L	2.59	2.47	0.12	2.54	123	2.48	0.01
6b	7a	S	3.05	3.07	−0.02	3.07	135	3.08	−0.01
6a	7b	L	2.51	2.42	0.09	2.49	143	2.43	0.00
6a	7a	L	2.48	2.37	0.11	2.44	102	2.37	0.00
2a	1a	S	2.51	2.60	−0.09	2.60	158	2.59	0.01
2a	1b	L	2.47	2.42	0.06	2.48	125	2.43	0.01
7b	6b	L	2.44	2.36	0.09	2.42	97	2.36	0.00
7b	6a	L	2.51	2.42	0.09	2.49	134	2.43	0.01
7b	15b	L	2.50	2.53	−0.03	2.60	153	2.52	0.01
7b	8	S	2.46	2.44	0.02	2.44	154	2.52	0.01
7b	14	S	2.61	2.54	0.08	2.54	131	2.54	0.01
7b	7a	S	1.76	1.76	0.00	1.76	98	1.76	0.00
MAE (%)				2.34		1.40		2.28	
St. dev. (%)				1.36		1.25		1.32	

NOESY and T-ROESY data used for calculation of τ_c (τ_m = 300 ms). Structure optimized to MPW1PW91/6-311 + g(d,p). H7b-H7a was used as reference. See text regarding $CDCl_3$ for further explanations

the MAE is essentially unchanged. It is obvious that the higher viscosity of DMSO leads to a longer τ_c and that the experimental data fit better to the calculated data. This may be contributed to the signal intensity being less dependent of τ_c for this rotational time regime.

The same analysis may be done by the two-field (2F) methods, to much the same conclusion. A comparison of the τ_c values from NOE/ROE and 2F analysis is found in Table 3.34, and the rotational correlation times are markedly different, which is problematic. In most cases the rotational correlation times are in the same general regime though, and the differences between corrected distances are negligible. This is currently under assessment in the group of Dr. Craig Butts and will not be further elaborated on.

Quinine

From a general assessment of the structure of quinine, it was postulated that the two rigid systems might experience different τ_c and that the inter-rigid NOEs would most certainly experience a τ_c different than that of the rigid NOEs, see Fig. 3.36 for clarification and the structure of quinine. Quinine was investigated in

Table 3.34 Progesterone in DMSO-d_6

H1	H2	τ_c (ps)	
		NOE/ROE	2F NOE
4	6a	85	88
17	16a	95	63
17	12a	105	97
17	14	138	130
6a	7b	143	90
6a	7a	102	165
2a	1a	158	124
2a	1b	125	119
7b	6b	97	255
7b	6a	134	165
7b	8	154	134
7b	7a	98	93

Comparison of experimental τ_c from NOE/ROE and two-field analysis (τ_m = 300 ms). NOESY and T-ROESY data used for calculation of τ_c at 500 or 600 MHz using a spinlock field of 5452 Hz with an offset at 5 ppm

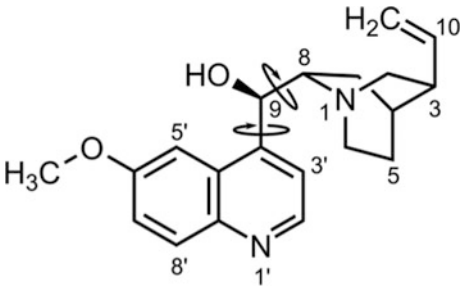


Fig. 3.36 The structure of quinine indicating the rotatable bonds connecting the rigid ring systems

chloroform and DMSO, but in this text only the chloroform data is analyzed, as the outcome and conclusions are very similar for data acquired in the two solvents.

First the rigid parts of quinine were investigated, Table 3.35, and there seem to be a slight difference in the rotational correlation time.

The fit, in the form of the MAE, is slightly improved by incorporating the τ_c term in the calculations, but again the distances are not generally improved, and the structural fit is already excellent.

The inter-rigid distances and their τ_c are found in Table 3.36. It should be noted that the populations were optimized in both instances from the structures discussed in Sect. 3.4.1, and that the populations only varied slightly.

Table 3.35 Quinine in $CDCl_3$

H1	H2	τ_c (ps)	Exp.	Exp. Corr.	Calc.	Impr.
2′	3′	188	2.48	2.45	2.49	−0.03
8′	7′	178	2.56	2.54	2.52	0.02
3′	2′	186	2.43	2.40	2.49	−0.03
3	2a	166	2.39	2.39	2.33	0.00
3	6b	162	3.13	3.14	3.32	0.01
3	4	149	2.53	2.55	2.49	−0.03
3	5b	156	2.48	2.49	2.44	−0.02
2a	2b	167	1.72	1.72	1.77	0.00
2a	3	163	2.32	2.33	2.33	0.01
8	7b	137	2.26	2.30	2.31	0.03
6b	6a	171	1.79	1.79	1.76	0.00
2b	2a	174	1.69	1.69	1.76	−0.01
2b	3	207	3.02	2.92	2.93	0.07
6b	5b	148	2.36	2.38	2.36	−0.02
6a	6b	169	1.78	1.78	1.76	0.00
6a	5a	143	2.29	2.32	2.35	0.03
MAE (%)			2.13	2.06		
St. dev (%)			1.37	1.42		

NOESY and T-ROESY data used for calculation of τ_c. Only rigid distances shown. Exp. Corr. are the distances corrected by the experimental τ_c values. Impr are the differences between experimental and theoretical distances for corrected minus uncorrected data; positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data. H6a-H6b was used as reference. Structures optimized to MPW1PW91/6-311 + g(d,p)

Table 3.36 Quinine in CDCl3

H1	H2	τ_c (ps)	Exp.	Exp. Corr.	Calc.	Impr.
8	3′	136	2.73	2.73	2.81	0.00
8	5′	178	2.39	2.39	2.51	0.00
8	9	180	2.55	2.55	2.65	0.00
9	3′	181	2.89	2.86	2.96	−0.03
9	5′	157	2.10	2.12	2.22	0.02
9	6a	145	2.53	2.56	2.43	−0.03
9	8	176	2.59	2.58	2.65	−0.01
9	7a	122	2.78	2.83	2.89	0.05
6a	9	143	2.45	2.48	2.43	−0.03
MAE (%)			3.35	3.49		
St. dev. (%)			1.33	1.13		

NOESY and T-ROESY data used for calculation of τ_c (τ_m = 300 ms). Only non-rigid distances shown and conformer populations were optimized before comparison. Exp. Corr. are the distances corrected by the experimental τ_c values. Impr are the differences between experimental and theoretical distances for corrected minus uncorrected data; positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data. H6a–H6b was used as reference

The τ_c was not markedly different for these flexible proton pairs than τ_c determined for the rigid parts, in contrast to our assumptions, and the MAE was actually slightly worsened by including the rotational correlation time in the calculations. When determining the populations, the difference between the NOE and NOE/T-ROE data sets are diminishing as illustrated in Fig. 3.37. Similarly the difference is minute in population determined from six or seven structures.

3.5.3 *Conclusion and Problems in Including Correlation Times*

While it would seem that including more data is always beneficial, the problem with the rotational correlation time is the erroneous methods that may be used to determine τ_c. The robustness of the NOE methods, due to the r^{-6} relationship, is thus reduced by introducing the correlation time, which does not benefit from a similar relationship. Due to generally low rotational correlation times in especially chloroform, any errors in NOESY and/or ROESY data will lead to large errors in the experimentally determined τ_c, as seen Fig. 3.38, due to only minor differences between the observables. As a consequence, an acceptable experimental error in the relative PANIC intensity of one to two will often be larger than the expected differences in PANIC intensities from NOESY/ROESY or two-field analysis experiments. For longer distances especially the introduction of the rotational correlation time into the equations is problematic, due to a larger influence of errors on the resulting distances.

The promise of more exact distances is present from these types of analyses. As seen in Fig. 3.39 the influence on the rotational correlation time on the distances obtained may be rather large, if the correlation time is sufficiently different. It is an enigma though, as the biggest difference from small rotational correlation time differences is in the regions where τ_c is hardest to measure correctly. In most cases it is thus not preferable to introduce the corrections into the equations for small molecules as this will only lead to more errors. It is also quickly evident that care

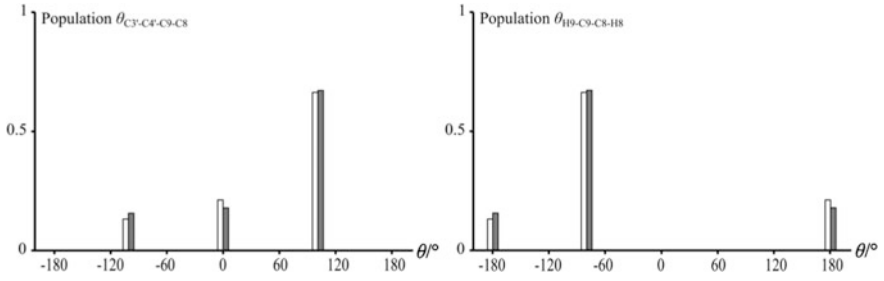


Fig. 3.37 The populations of quinine conformers from NOE (white) and NOE/T-ROE (grey) distances using the three essential conformers identified in Sect. 3.4.1

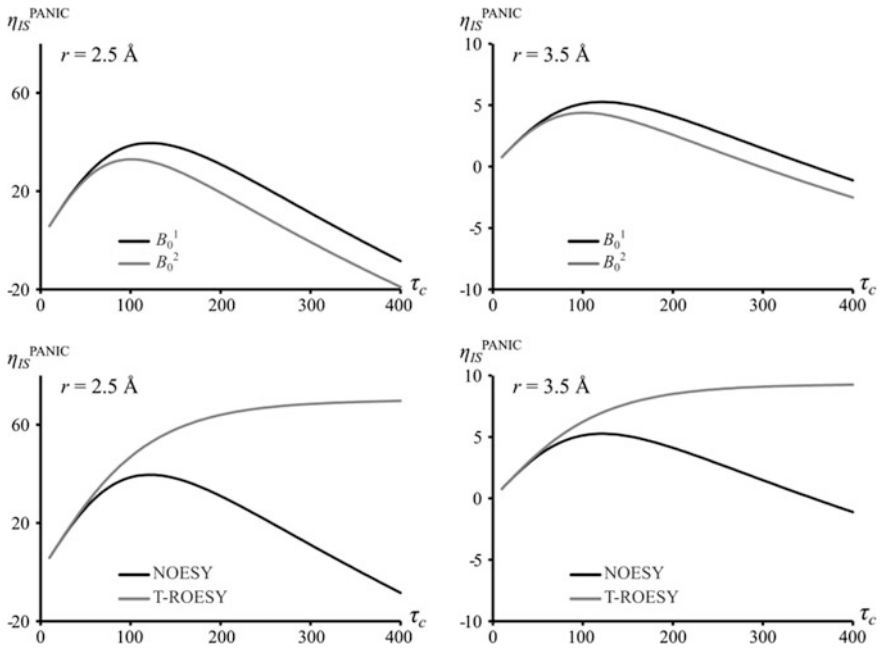


Fig. 3.38 Theoretical PANIC intensities from NOESY and T-ROESY or two-field NOESY analyses. Calculated from (3.8) and (3.11) using ω = 500 MHz (NOE/T-ROE), or ω_1= 500 MHz $\left(B_0^1\right)$ and ω_2= 600 MHz $\left(B_0^2\right)$. r given in the figures. A spinlock field of 5452 Hz with an offset at 5 ppm was assumed for the T-ROESY calculations

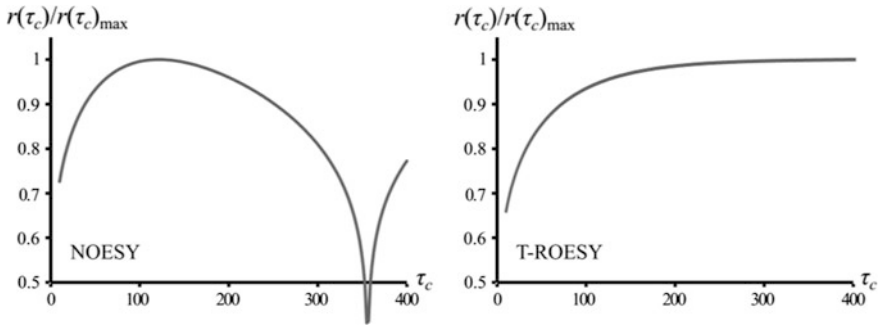


Fig. 3.39 Theoretical distance correction factors as a fucntion of the rotational correlation time (in ms) for NOESY and T-ROESY. Calculated from (3.8) and (3.11) using ω = 500 MHz, and a spinlock field of 5452 Hz with an offset at 5 ppm

should be taken in the NOESY crossover region, and that the ROESY experiment is not just superior in the relative intensity, but also in a minimal dependence on the rotational correlation time in the crossover region (these two observations are coupled but maybe not always realized). Of course ROESY type experiments have

their own challenges in this regard and may be in need of correcting before being used quantitatively.

In conclusion the attempt to obtain experimental distances with a higher correlation to those of 3D structures was not successful. While the methods may be used to estimate τ_c and check the assumption of roughly equal τ_c between proton pairs, direct usage of τ_c in distance calculations is not recommended.

3.6 Experimental

General on structural calculations

The theory used for obtaining distances from NOESY and ROESY experiments is outlined in the theory section (Sect. 2.1). A build up curve was always established using (25), 50, 100, 150 and 200 (and 300) ms mixing times for each cross-peak and each curve were evaluated. Standard setups for all NMR experiments and the spectrometers used are found in Appendix A1.

All *J*-coupling constants were extracted from 1D ^{1}H spectra, or DQF-COSY spectra if necessary due to overlapping resonances.

Fungal extraction
Solvents used

The solvents used were all HPLC grade. The water used was milliQ water from the milliQ Water Purification system by Merck Millipore.

Media used

The media used was yeast extract sucrose agar (YES) obtained from the stock at the Department of Systems Biology.

Fungi

Aspergillus homomorphus, IBT 21893 (or CBS 101889), originating from soil in the area of the Dead Sea in Israel. Inoculated by three point stabs. All plates were kept at 30 °C in darkness after inoculation for 7 days.

Extraction

Fungi were collected in double bagged stomacher bags; ten plates to a bag. 150 mL EtOAc (1% FA) was added and the content processed in a Stomacher for 30–60 s. After one hour the extract was decanted into vials or flasks through a filter, the solvent was removed using N_2 and rotary evaporators. The bags were refilled with 150 mL EtOAc (1% FA) and left over night before being re-collected as described.

The precipitate from the pooled extracts were dissolved in 200 mL 9:1 MeOH: H_2O and extracted with 180 mL heptane. The MeOH/H_2O phase was added 160 mL H_2O and extracted 5 times with 100 mL DCM. All phases were dried using N_2 and rotary evaporators.

Isolera One

An Isolera One flash chromatography system by Biotage with auto-fractionation was used for fractioning. The column was a 120 g Biotage Snap KP-C18-HS with a column volume of 132 mL and fractions of 115 mL (max) were collected, with a flow speed of 40 mL/min.

Waters

A semi-preparative HPLC was used for further purification; the Waters 600 Controller coupled to a Waters 996 Photodiode Array Detector by Waters with a flow rate of 5 mL/min, using a Luna II C18, 5 µm, 250 × 10 mm column by Phenomenex. The runs are found Table 3.37. The samples were dissolved in MeOH before injection.

NMR—natural compounds

All compounds were dissolved in DMSO-d_6 in 3 mm (180 µL) or 5 mm (500 µL) NMR tubes. Standard pulse sequences used for structural elucidation, at either 500 or 800 MHz as described in Appendix A1. 2D ROESY spectra were used for distance calculations of homomorphosin A (23 mM) using τ_m = 200 ms and 2D NOESY spectra for cyclomorphosin A (15 mM), τ_m = 200 ms. For epi-10,23-dihydro-24,25-dehydroaflavine (43 mM) 2D NOESY spectra were used, τ_m = 150 ms. The remaining natural products, homomorphosterol (12 mM), aculene A (19 mM), aculene B (23 mM) and aculene C (18 mM) were solved from 1D ^{1}H and DQF-COSY spectra.

Simulations

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations [74], using the program MacroModel version 10.8 [75, 76]. The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using torsional sampling with an energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization [75, 77]. The method will alter the dihedral angle of bonded nuclei by random Monte Carlo simulations, before it minimizes the new structure and, if it is indeed new as defined by relative atom position and within a given energy of the minimum, saves the structure. Also low-mode was used in certain simulations when stated in the text. Low mode sampling is based on an assumption that low-frequency vibrational mode

Table 3.37 Waters runs

Run fraction	Column	Gradient MeCN[a]/H_2O[a]	Run time (min)	Fraction name	Name	Amount isolated (mg)
CHO012206	Luna II	30–40% MeCN	17	CHO013409	CM-B	1.1
CHO013406	Luna II	20–60% MeCN	20	CHO013801	CM-A	2.3

a: +50 ppm TFA. b: isocratic run

eigenvectors connect energy minima by low-energy paths on the conformational energy surface. Thus by moving coordinates along a path given by the low-frequency modes, large energy barriers between energy minima may be traversed. These minima might not be identified by other methods due the large energy barriers connecting the minima. If moving along the low-mode eigenvector results in a new energy well, energy minimization is performed as for any other conformational sampling [78].

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations [79]. Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated. NMR shielding tensors were calculated by MPW1PW91/6-311 + G(2d,p) using GIAO and the PCM-SCRF model for implicit solvent [30].

J_{HH} were calculated by B3LYP/6-31 g(d,p) u + 1 s considering only the Fermi Contact term. This is practically done by inserting the following in the Gaussian input file: nmr = fconly iop(3/10 = 1100000).

Rigid to semi-rigid

For rigid compounds which may be described by one to ten distinct conformers, e.g. quinine, DFT optimized structures are used as the basis of distance and *J*-coupling constant calculations. This is feasible due to the small amount of conformers, and the higher level of theory used should lead to more appropriate bond lengths, angles etc. This is equal to the approach utilized with much success in the group of Craig Butts, whom I visited on my external stay [80–83]. In general the populations of the conformers may here be described in detail from an iterative fitting of data from conformer populations to the experimental data. If multiple structures were used (e.g. quinine) the Excel Solver was used for conformer population analysis, using the evolutionary solving method. The reference distance was here allowed to differ by up to 5% to fit the data the best, except for the one structure data of homomorphosin A and Epi-10,23-dihydro-24,25-dehydroaflavine.

Flexible

For more flexible compounds, where the amount of conformers is large and/or ill described, an approach based in DFT calculations is not feasible. One could use centroids and optimize these, but a lot of subtleties are lost, and the choice of centroids will introduce a bias into the data. The approach used was taken out of necessity, but the robustness of the approach makes it highly favorable. It was realized that the systems could not be described from a few representative conformers, and the approach instead compares the NOE/*J* data to a large amount of conformers from MD or MM simulations. The conformers are fitted iteratively to give the best correlation between data from the averaged conformers to the experimental data with no regards to the relative energy or other properties of the conformers. This means that the output conformations should be seen as representative of the conformational space inhabited by the compound. While the approach will by design be less discriminative in e.g. determining relative stereochemistry, due to the disregard of relative energy, it is less prone to

misinterpretation due to faulty energy determination. Like all approaches it does rely on the different force field methods to output meaningful conformers, but by disregarding the energy of these, it is more likely that conformational space is covered. The cutoff of 50 kJ/mol was chosen for standard MM simulations to hopefully cover all of conformational space, and the NMR data was then used to gain information of the actual conformational space. The high cutoff is used to diminish the influence of errors in the calculations. Since the NOE data is dominated by short distances, the average distances observed may be changed dramatically by short distances in lowly populated conformers, and thus excluding these may severely influence the conclusions.

The comparison of experimental data to that of 3D structures was performed by an in house written Matlab® script which iteratively minimizes the MAE using 10,000 steps. In the structural evaluation and iterative identification of the structures which gave the highest correlation to the NOE and *J*-coupling constants, both datasets were fitted independently and mixed, by weighing the differences as $\Delta = \Delta(\text{NOE}) + 1/10\Delta(J)$. The latter data is shown throughout. Thus the unconstrained NMR structures were determined by averaging only the conformations that contributed to a best overall correlation to the data, by the lowest RMSD in both distances and coupling constants. The reference distance was here always equal to that of the averaged structures. The averaging was performed as described in Sects. 2.1.9 and 2.2.1.

Marfey's reagent

100 µg of homomorphosin A, cyclomorphosin A or aculene A and B were hydrolyzed with 200 µL 6 M HCl in a 2 mL vial at 110 °C for approximately 24 h. The solvent was removed at N_2 evaporator. To the vials, and vials with 50 µL (2.5 µmol) L- or D-amino acids (valine or proline), were added 50 µL H_2O, 20 µL 1 M $NaHCO_3$ and 100 µL 1% 1-fluoro-2,4-dinitrophenyl-5-L-alanineamide (FDAA) in acetone. The vials were left at 40 °C for 1 h. The solvent was removed at N_2 evaporator, MeOH added and HPLC-DAD-MS data acquired.

Synthetic compounds

The synthetic compounds were supplied by the group of Prof. Christian A. Olsen (azumamides) or Prof. Morten Meldal (molecular recognition).

Azumamides

The azumamide analogues were dissolved in DMSO-d_6 in 3 mm tubes (180 µL). Spectra were acquired at 500 and 800 MHz using 2D NOESY (500 and 800 MHz) and 2D ROESY (500 MHz) sequences, τ_m = 150 ms. 2D ROESY spectra at 500 MHz were used due to a superior *S/N*. 4096 data points were recorded in the direct and 512 in the indirect dimension. Relaxation delays were from 2 to 4 s, tested up to 7 s (>5 × T_1). Most nuclei were determined to have T_1 values up to approximately 1–1.3 s relaxation times. All *J*-couplings were extracted from 1D ^{1}H or DQF-COSY spectra, and *J*-couplings from dihedral angles were calculated by the Karplus equation or the HLA equation for the β-peptide protons.

Constrained simulations in implicit DMSO were conducted in MacroModel using the force field MMFFs or in implicit water using OPLS2005. DMSO was treated as a dielectric constant of 47.0. A conformational search was performed by the program MacroModel using torsional/low-mode sampling with an energy cutoff of 30 kJ/mol, 30,000 steps and CPRG minimization [75, 77]. The distances from NOE intensities were applied to the structure and allowed to differ 20% governed by a force constant of 100 kJ/Å2.

Molecular dynamics simulations of azumamide A, desmethyl-azumamide C and epi-azumamide E were performed by the program Desmond, carried out by Post. Doc. Niels Christensen [76]. Individual structures were solvated in a TIP3P or DMSO box using a 10 Å buffer, and neutralized by a single Na^+ counter ion. The default minimization protocol in Desmond was used, consisting of minimization with solute constraints, minimization without constraints, Berendsen NVT simulation at T = 10 K with small time steps and constraints on heavy solute atoms, Berendsen NPT simulation at T = 10 K with constraints on heavy solute atoms, Berendsen NPT simulation with constraints on heavy solute atoms, and unrestrained Berendsen NPT simulation. In the MD production runs, 250 ns NPT simulations with periodic boundary conditions were run for each system. The Nose-Hoover, 8 chain thermostat was used to regulate temperature to 300 K with a relaxation time of 1.0 ps. The Martyna-Tobias-Klein barostat was used to regulate pressure to 1 bar with isotropic coupling and a relaxation time of 2.0 ps. Equations of motion were integrated using the RESPA10 integrator with bonded (2.0 fs), near (2.0 fs), and far (6.0 fs) time steps. Non-bonded interactions were subjected to a 9 Å cut-off. Long-range electrostatics were treated with the smooth-particle mesh Ewald method. All atomic coordinates for the simulations were saved at 10 ps intervals.

Molecular recognition

The Target 20 (33 mM) and Ligand 7 (34 mM) were dissolved in H_2O/10% D_2O in 5 mm tubes (0.5 mL). The pH was adjusted by NaOH (0.1 M). Spectra were acquired at 800 MHz using 2D NOESY and 2D ROESY sequences using water suppression using either presaturation or watergate. Distances obtained from both types of water suppressions were compared and the differences were negligible. 2D NOESY spectra using presaturation were used, τ_m = 200 ms. Prior to acquisition N_2 was bubbled through the sample to remove O_2. The solutions were mixed, 0.5 mL transferred to a NMR tube and the acquisitions repeated as for the isolated peptides.

Titrations of ligand 7 with target 20 were performed at pH 5.5 and 6.5. Both solutions of the peptides were adjusted to the target pH prior to titrations, which were performed in 3 mm NMR tubes. 0.9 (pH 5.5) or 0.7 (pH 6.5) mg of Ligand 7 was dissolved in 180 μL 10% D_2O/H_2O and adjusted to the target pH with 0.1 M NaOH and double the molar amount of Target 20 was dissolved in 100 μL 10% D_2O/H_2O and pH adjusted with 0.1 M NaOH. Titrations were then performed stepwise by adding the amount stated (e.g. 10% equals to 4.8 μL Target solution added) in the figures, and 1D ^1H spectra were acquired with 128 scans.

The structures, Target 20 and Ligand 7, were built in MOE and both subjected to constrained dynamics in water from NOE derived distances by Prof. Morten Meldal. Starting from several rounds of 3 ns annealing from 700 to 300 K, using AMBER 12 force field, the resulting structures were simulated at 300 K for 5 ns and energy minimized. The NOE-constraints from the complex were used for the initially as the set of distances were larger and presented better *S*/*N* than those of the individual compounds. The individual structures were manually interacted assuming interactions of arginines and carboxylates, and the resulting complexes were subjected to molecular dynamics. Two orientations, out of six attempted, gave no significant violation of NOE's or change of structure. These two orientations were subjected to extended molecular dynamics for 5 ns with the constraints maintained. One orientations lead to productive binding and surface complementarity and gave a significantly lower calculated energy compared to the other. In this structure all backbone distances constraints determined from relative NOE intensities were accommodated. The complex was optimized again without NOE constraints, and only minor adjustments of the two structures were observed.

NOE constraints from individual Target 20 and Ligand 7 experiments, were employed to structures from the optimized complex, which were then subjected to annealing from 700 to 300 K. The Target 20 was essentially unchanged the Ligand 7 changed to a partially opened.

Other compounds

Spectra of strychnine (124 mM), progesterone (127 mM) and quinine (123 mM) were all acquired in 5 mm tubes (0.5 mL), dissolved in the solvents given in the text (concentration in $CDCl_3$, similar in DMSO-d_6). 1D NOESY and 1D ROESY experiments were used and only resonances with good baseline separation were used as irradiated peaks, τ_m = 300 ms. For the ROESY experiments a spinlock field of 5452 Hz with an offset at 5 ppm was used at 500 MHz. For two-field analyses the other field was 600 MHz. All structures were optimized to a MPW1PW91/6-311 + g(d,p) level of theory.

References

1. P.M. Dewick, *Medicinal Natural Products: A Biosynthetic Approach*, 3rd edn. (Wiley, Hoboken, 2009)
2. N.P. Keller, G. Turner, J.W. Bennett, Nat. Rev. Microbiol. **3**, 937–947 (2005)
3. F. Kempken, M. Rohlfs, Fungal Ecol. **3**, 107–114 (2010)
4. A.M. Calvo, R.A. Wilson, J.W. Bok, N.P. Keller, Microbiol. Mol. Biol. Rev. **66**, 447–459 (2002)
5. D.J. Newman, G.M. Cragg, J. Nat. Prod. **79**, 629–661 (2016)
6. P.D. Leeson, B. Springthorpe, Nat. Rev. Drug Discov. **6**, 881–890 (2007)
7. J.L. Meier, M.D. Burkart, Chem. Soc. Rev. **38**, 2012–2045 (2009)
8. M.A. Fischbach, C.T. Walsh, Chem. Rev. **106**, 3468–3496 (2006)
9. T. Weber, H.U. Kim, Synth. Syst. Biotechnol. **1**, 69–79 (2016)
10. J. Varga, J.C. Frisvad, S. Kocsubé, B. Brankovics, B. Tóth, G. Szigeti, R.A. Samson, Stud. Mycol. **69**, 1–17 (2011)

11. K.F. Nielsen, J.M. Mogensen, M. Johansen, T.O. Larsen, J.C. Frisvad, Anal. Bioanal. Chem. **395**, 1225–1242 (2009)
12. R. Steiman, P. Guiraud, L. Sage, F. Seigle-Murandi, Syst. Appl. Microbiol. **17**, 620–624 (1995)
13. M.L. Chiotta, A. Susca, G. Stea, G. Mulè, G. Perrone, A. Logrieco, S.N. Chulze, Int. J. Food Microbiol. **149**, 171–176 (2011)
14. R.A. Samson, P. Noonim, M. Meijer, J. Houbraken, J.C. Frisvad, J. Varga, Stud. Mycol. **59**, 129–145 (2007)
15. J.M. Berg, J.L. Tymoczko, L. Stryer, *Biochemistry*, 6th edn. (W. H. Freeman and Company, New York, 2007)
16. Y.N. Han, K.H. Hwang, B.H. Han, Arch. Pharm. Res. **28**, 159–163 (2005)
17. K.H. Hwang, Y.N. Han, B.H. Han, Arch. Pharm. Res. **24**, 202–206 (2001)
18. A.F. Morel, E.C. Machado, L.A. Wessjohann, Phytochemistry **39**, 431–434 (1995)
19. M.A. Mostardeiro, V. Ilha, J. Dahmer, M.S.B. Caro, I.I. Dalcol, U.F. da Silva, A.F. Morel, J. Nat. Prod. **76**, 1343–1350 (2013)
20. M.H. Park, D.-Y. Suh, B.H. Han, Phytochemistry **43**, 701–704 (1996)
21. C.T. Walsh, R.V. O'Brien, C. Khosla, Angew. Chem. Int. Ed. Engl. **52**, 7098–7124 (2015)
22. M. Köck, G. Schmidt, I.B. Seiple, P.S. Baran, J. Nat. Prod. **75**, 127–130 (2012)
23. R. Bhushan, H. Brückner, Amino Acids **27**, 231–247 (2004)
24. P. Marfey, Carlsberg Res. Commun. **49**, 591–596 (1984)
25. W.K. Wilson, R.M. Sumpter, J.J. Warren, P.S. Rogers, B. Ruan, G.J. Schroepfer Jr., J. Lipid Res. **37**, 1529–1555 (1996)
26. S.G. Smith, J.M. Goodman, J. Am. Chem. Soc. **132**, 12946–12959 (2010)
27. L.W. Parks, W.M. Casey, Annu. Rev. Microbiol. **49**, 95–116 (1995)
28. Y. Zhang, X.-M. Li, P. Proksch, B.-G. Wang, Steroids **72**, 723–727 (2007)
29. R. Jain, T. Bally, P.R. Rablen, J. Org. Chem. **74**, 4017–4023 (2009)
30. M.W. Lodewyk, M.R. Siebert, D.J. Tantillo, Chem. Rev. **112**, 1839–1862 (2012)
31. L.M. Petersen, C. Hoeck, J.C. Frisvad, C.H. Gotfredsen, T.O. Larsen, Molecules **19**, 10898–10921 (2014)
32. M.R. TePaske, J.B. Gloer, Tetrahedron **45**, 4961–4968 (1989)
33. J.B. Gloer, M.R. Tepaske, J.S. Sima, D.T. Wicklow, P.F. Dowd, J. Org. Chem. **53**, 5457–5460 (1988)
34. A.R. Maolanon, J.S. Villadsen, N.J. Christensen, C. Hoeck, T. Friis, P. Harris, C.H. Gotfredsen, P. Fristrup, C.A. Olsen, J. Med. Chem. **57**, 9644–9657 (2014)
35. M. Li, C. Hoeck, S. Schoffelen, C.H. Gotfredsen, M. Meldal, Chem. Eur. J. **22**, 7206–7214 (2016)
36. Y. Nakao, S. Yoshida, S. Matsunaga, N. Shindoh, Y. Terada, K. Nagai, J.K. Yamashita, A. Ganesan, R.W.M. Van Soest, N. Fusetani, Angew. Chemie **118**, 7715–7719 (2006)
37. P. Bertrand, Eur. J. Med. Chem. **45**, 2095–2116 (2010)
38. A.J.M. de Ruijter, A.H. van Gennip, H.N. Caron, S. Kemp, A.B.P. van Kuilenburg, Biochem. J. **370**, 737–749 (2003)
39. I. Izzo, N. Maulucci, G. Bifulco, F. De Riccardis, Angew. Chemie **118**, 7719–7722 (2006)
40. N. Maulucci, M.G. Chini, S. Di, I. Izzo, E. Cafaro, A. Russo, P. Gallinari, C. Paolini, M.C. Nardi, A. Casapullo, R. Riccio, G. Bifulco, F. De Riccardis, J. Am. Chem. Soc. **129**, 3007–3012 (2007)
41. B. Vögeli, P. Güntert, R. Riek, Mol. Phys. **111**, 437–454 (2013)
42. Y. Levy, N. Onuchic, Annu. Rev. Biophys. Biomol. Struct. **35**, 389–415 (2006)
43. G.R. Crabtree, N.A. Clipstone, Annu. Rev. Biochem. **63**, 1045–1083 (1994)
44. E.J. Sundberg, R.A. Mariuzza, Adv. Protein Chem. **61**, 119–160 (2002)
45. R.E. Babine, S.L. Bender, Chem. Rev. **97**, 1359–1472 (1997)
46. M. Renil, M. Ferreras, J.M. Delaisse, N.T. Foged, M. Meldal, J. Pept. Sci. **4**, 195–210 (1998)
47. M. Meldal, B. Wu, F. Diness, R. Michael, G. Hagel, ChemBioChem **12**, 2463–2470 (2011)
48. E.P. Diamandis, T.K. Christopoulos, Clin. Chem. **37**, 625–636 (1991)
49. B.W. Sigurskjolds, D.R. Bundle, J. Biol. Chem. **267**, 8371–8376 (1992)

50. H. Kaur, M. Datt, M.K. Ekka, M. Mittal, A.K. Singh, S. Kumaran, Biochimie **93**, 175–186 (2011)
51. S. Sasaki, M. Takagi, Y. Tanaka, M. Maeda, Tetrahedron Lett. **37**, 85–88 (1996)
52. T. Young, R. Abel, B. Kim, B.J. Berne, R.A. Friesner, Proc. Natl. Acad. Sci. U.S.A. **104**, 808–813 (2007)
53. S. Macura, B.T. Farmer II, L.R. Brown, J. Magn. Reson. **70**, 493–499 (1986)
54. P. Thordarson, Chem. Soc. Rev. **40**, 1305–1323 (2011)
55. J.N.S. Evans, *Biomolecular NMR Spectroscopy*, 1st edn. (Oxford University Press, Oxford, 1995)
56. T. Burgi, A. Baiker, J. Am. Chem. Soc. **120**, 12920–12926 (1998)
57. J.T. Kowalik, T. Lipinska, B.J. Oleksyn, J. Sliwinski, Enantiomer **4**, 389–410 (1999)
58. G.D.H. Dijkstra, R.M. Kellogg, H. Wynberg, J. Org. Chem. **55**, 6121–6131 (1990)
59. E.M.O. Yeboah, S.O. Yeboah, G.S. Singh, Tetrahedron **67**, 1725–1762 (2011)
60. T.H.A. Silva, A.B. Oliveira, W.B. De Almeida, Struct. Chem. **8**, 95–107 (1997)
61. A. Sen, A. Bouchet, V. Lepe, K. Le Barbu-debus, D. Scuderi, F. Piuzzi, A. Zehnacker-Rentien, J. Phys. Chem. A **116**, 8334–8344 (2012)
62. A. Urakawa, D.M. Meier, H. Rüegger, A. Baiker, J. Phys. Chem. A **112**, 7250–7255 (2008)
63. S.V. Efimov, I.A. Khodov, E.L. Ratkova, M.G. Kiselev, S. Berger, V.V. Klochkov, J. Mol. Struct. **1104**, 63–69 (2016)
64. R. Pendrill, E. Säwén, G. Widmalm, J. Phys. Chem. B **117**, 14709–14722 (2013)
65. H. Kessler, C. Griesinger, R. Kerssebaum, K. Wagner, R.R. Ernst, J. Am. Chem. Soc. **109**, 607–609 (1987)
66. M. Mackeen, A. Almond, I. Cumpstey, S.C. Enis, E. Kupce, T.D. Butters, A.J. Fairbanks, R. A. Dwek, M.R. Wormald, Org. Biomol. Chem. **4**, 2241–2246 (2006)
67. A. Kjellberg, G. Widmalm, Biopolymers **50**, 391–399 (1999)
68. X. Mao, T. Zhang, M. Baur, H. Kessler, J. Chem. Phys. **111**, 8253–8254 (1999)
69. D. Neuhaus, M. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 1st edn. (VCH Publishers, Weinheim, 1989)
70. D. Lee, C. Hilty, G. Wider, K. Wüthrich, J. Magn. Reson. **178**, 72–76 (2006)
71. T.-L. Hwang, M. Kadkhodaei, A. Mohebbi, A.J. Shaka, Magn. Reson. Chem. **30**, S24–S34 (1992)
72. T.-L. Hwang, A.J. Shaka, J. Am. Chem. Soc. **114**, 3157–3159 (1992)
73. U. Olsson, E. Säwën, R. Stenutz, G. Widmalm, Chem. Eur. J. **15**, 8886–8894 (2009)
74. Maestro (Schrödinger, LLC, New York, 2015)
75. Schrödinger (LLC, New York, NY 2016)
76. Desmond (D. E. Shaw Research, New York, 2015)
77. G. Chang, W.C. Guida, W.C. Still, J. Am. Chem. Soc. **111**, 4379–4386 (1989)
78. I. Chen, N. Foloppe, Bioorg. Med. Chem. **21**, 7898–7920 (2013)
79. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K. N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V. G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in *Gaussian 09, Revision B.01* (Gaussian Inc., Wallingford, CT, 2010)
80. C.R. Jones, C.P. Butts, J.N. Harvey, Beilstein J. Org. Chem. **7**, 145–150 (2011)
81. C.P. Butts, C.R. Jones, Z. Song, T.J. Simpson, Chem. Commun. **48**, 9023–9025 (2012)
82. C.P. Butts, C.R. Jones, E.C. Towers, J.L. Flynn, L. Appleby, N.J. Barron, Org. Biomol. Chem. **9**, 177–184 (2011)
83. C.P. Butts, C.R. Jones, J.N. Harvey, Chem. Commun. **47**, 1193–1195 (2011)

Chapter 4
Development of NMR Experiments for Determination of Long-Range *J*-Coupling Constants

The experimental data presented in this chapter concerns the work on novel NMR experiments to extract coupling constants from small organic molecules, and is based on two articles [1, 2]. Work on the first article was performed in collaboration with Ole W. Sørensen, Louise Kjaerulff and Andrew J. Benie, and work for the second article was performed in collaboration with Ole W. Sørensen. Louise Kjærulff presented the work of combining the five pulse sequences in the original S^3 HMBC, *vide infra*, in details in her Ph.D. thesis "NMR structural studies of oligosaccharides and other natural products" [3], and the topics will be mostly outlined in this text.

4.1 3D Structural Information from Long-Range *J*-Coupling Constants

The most commonly used 3D structural information obtained from *J*-coupling constants is based on data from $^3J_{HH}$, which was discussed in Sect. 2.2.1 where general scalar coupling theory is also found. The focus on $^3J_{HH}$-coupling constants is mainly due to the ease of extraction of the coupling constants and the availability of a plethora of Karplus- or Karplus-like equations, which makes dihedral information readily obtainable [4–6]. While the more generalized Karplus-like equations may be less accurate than e.g. DFT, the ease of calculation is a major benefit, and the scope of use is continually extended by specific equations for specific tasks [7].

For long-range *J*-coupling constants, defined as $^nJ_{XH}$, with $n > 3$ for ^1H and $n > 1$ for ^{13}C, the coupling constants are not generally easily obtained nor easily correlated to a 3D structural parameter. While some Karplus equations have been developed for $^2J_{CH}$ and $^3J_{CH}$ systems, they are generally very specific and used in only a few fields, perhaps due to difficulties in accessing the coupling constants or lack of general applicability of the equations [8–11]. It has been suggested to utilize

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_4

a strong/medium/weak coupling constant approach, in spiritual resemblance to NOE work for macromolecules [8]. This approach is a viable option for rigid structures, but difficulties may be envisioned from an increase in flexibility, even though some examples of flexible structures were presented [8]. For other long-range coupling constants ($^{n}J_{XH}$, $n > 3$, X = H,C) Karplus equations are complicated by addition of extra bond and thus extra dihedral angles to the equations, though very specific equations were empirically derived for specific cases of $^{4}J_{HH}$-coupling constants [12, 13].

The data from the long-range coupling constants is thus usually compared to more time-consuming theoretical DFT calculations to gain structural information [14–16]. While this may limit the amount of conformers included in theoretical data calculations, access to increasingly larger computers decreases the calculation time and has made the information increasingly accessible. Since any increase in information of structural parameters will inevitably lead to more structural knowledge, the information from long-range coupling constants holds considerable value and should not be discarded [17].

4.1.1 Published Methods to Extract Long-Range Coupling Constants

This section will serve as a brief introduction to the currently available NMR experiments which may be used for the extraction of long-range coupling constants. Regarding the extraction of long-range homonuclear coupling constants, E.COSY type experiments or standard 1D ^{1}H or DQF-COSY sequences may be used by simulation of coupling patterns or displacement of resonances [18–20]. In the DQF-COSY case the measurements described were performed on rigid structures with larger long-range coupling constants [20].

In comparison, a plethora of experiments has been developed for the extraction of heteronuclear coupling constants, but the NMR community has still to decide on a golden standard, akin to the 1D ^{1}H spectra for homonuclear coupling constants [21, 22]. Characteristics for a golden standard is, per Parella and Espinosa [21], an experiment which displays good sensitivity and high resolution, is generally applicable, accurate, sign-selective and simple to run and analyze. This experiment could of course be the coupled 1D ^{13}C spectrum, but here the analysis is cumbersome and NMR simulations are usually necessary due to the complexity of the multiplets [23–25]. The lack of a reference experiment, in addition to the differences between heteronuclear coupling constants determined from the available experiments, may in principle lead to overfitting of data. This will be discussed *vide infra*.

The experiments for determining long-range heteronuclear coupling constants may be, coarsely, divided into two categories, as being either COSY/TOCSY or purely HMBC/HSQMBC based [21]. Note that the experiments in TOCSY

category of course also utilize long-range heteronuclear coupling constants and may thus also be perceived as e.g. HSQMBC based. A schematic representation between the TOCSY and HMBC/HSQMBC based coupling topologies is found in Fig. 4.1.

The NMR available experiments vary with respect to how the *J*-coupling constants are extracted, from *J*-quantitative (comparison of cross-peak volume) to peak-splittings measured as in- or antiphase (IP or AP), IPAP, *J*-resolved, E.COSY type and TROSY patterns, exemplified in Fig. 4.2 [21].

The determination of the coupling constants from the coupling patterns, range from integration, direct measurements of peak-splitting and displacement of subspectra to more or less complicated theoretical simulations of multiplets [21]. For experiments where the *J*-coupling constants are measured in the F1 (indirect) dimension, the resolution is determined by number of increments and spectral width, which will often lead to excessively long experiment times. To alleviate this problem, the *J*-coupling constants are usually scaled by a factor *k*, which may be chosen according to the minimum required discrimination of observable *J*-coupling constants. For example; if coupling constants of approximately 1 Hz need to be determined and $k = 15$, the FID resolution in F1 needs to be 15 Hz per point. The size of *k* is restricted by the fact that the coupling evolution time increases with *k*,

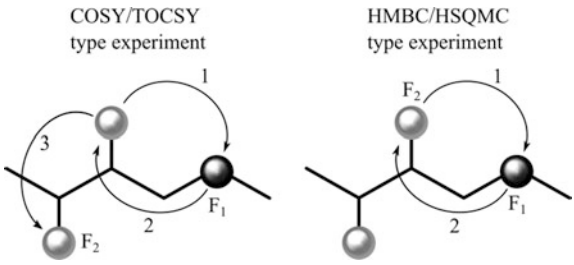


Fig. 4.1 The *J*-coupling transfer mechanism leading to spin correlations of COSY/TOCSY and HMBC/HSQMBC type experiments. The observed coupling constants are indicated by the observed nuclei in the F1 and F2 dimension. The COSY/TOCSY type utilizes an $^{n}J_{CH}$-coupling to build a correlation which is propagated to neighboring protons from $^{n}J_{HH}$-coupling. The HMBC/HSQMBC type utilizes only the $^{n}J_{CH}$-coupling. Grey: ^{1}H, black ^{13}C. Inspired from literature [21]

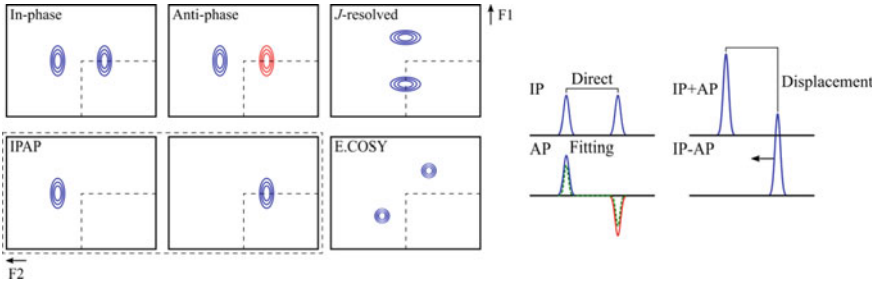


Fig. 4.2 Illustration of cross-peaks and cross sections from the different methods for the extraction of coupling constants. Inspired from literature [21]

and thus large values may be problematic due to T_2 relaxation of the nuclei and concurrent losses in sensitivity.

Some of the published experiments are given in Table 4.1 to exemplify the peak patterns, measurement method, and general considerations. Advantages and disadvantages of the presented experiments will differ depending of the information needed, and are to some extent subjective.

The presented list is not complete, nor are the given comments, and multiple experiments not mentioned here are available in the literature and e.g. various reviews include additional examples [21, 40]. In the following sections, NMR experiments that provide easy access to $^{n}J_{HH}$ as well as $^{n}J_{CH}$ including the relative sign of the coupling constants will be described.

Table 4.1 Examples of experiments for the determination of long-range coupling constants

Name	Peak pattern	Type	Comments	References
Quantitative HMBC	*J*-quantitative	$^{n}J_{CH}$	Compares the HMBC peaks by volume and calculation is needed Not sign-selective	[26, 27]
EXSIDE	*J*-resolved	$^{n}J_{CH}$	Not sign-selective F1 dimension limit resolution Selective version exists to e.g. increase resolution	[28, 29]
XLOC	E.COSY	$^{n}J_{HH}$	HH in F2 dimension Not sign-selective	[30]
J-HMBC	*J*-resolved	$^{n}J_{CH}$	Not sign-selective F1 dimension limit resolution. *J*-scaling is used	[30]
HECADE	E.COSY	$^{n}J_{CH}$	CH in F2 dimension Sign-selective	[31]
P.E.HSQMBC	E.COSY	$^{n}J_{CH}$ $^{n}J_{HH}$	CH and HH coupling constants Sign-selective for HH. Not sign-selective for CH Complicated patterns CH in F1 dimension. *J*-scaling is used	[32]
HSQC-TOCSY	In-phase/IPAP	$^{n}J_{CH}$	CH in F2 dimension Sign-selective version exists	[33–35]
HSQMBC	In-phase/anti-phase/IPAP	$^{n}J_{CH}$	CH in F2 dimension Not sign-selective	[36, 37]
Phase sensitive HMBC	Anti-phase	$^{n}J_{CH}$	CH in F2 dimension Not sign-selective Simulations of cross-peaks needed	[38]
HSQMBC-COSY/TOCSY	IPAP	$^{n}J_{CH}$	Sign-selective CH in F2 dimension Selective versions are available	[39]

Peak pattern (see Fig. 4.2 for examples), type and references are given. Note that several experiments may also be used for determination of $^{n}J_{NH}$ as well as $^{n}J_{CH}$-coupling constants. The table is inspired by Parella et al. [21]

4.2 S^3 HMBC *Homo*

The spin-state selective (S^3) HMBC *homo* (homonuclear in regards to the determined *J*-coupling constants) pulse sequence was developed in order to measure long-range homonuclear coupling constants utilizing the S^3 or E.COSY methodology of sub-spectra displacement [1, 41]. The practical determination of coupling constants is thus in line with the IPAP method described in the previous section [21]. The subspectra are linear combinations of spectra of a standard HMBC pulse sequence and a second pulse sequence with an additional $\pi/2$ S_z rotation under the one bond coupling evolution. The selected multiplicity edited pulse sequences are the improved multiplicity edited HMBC as the standard (a–c) and the HAT HMBC as the second (d, e) pulse sequence, see Fig. 4.5 and Eq. (4.1) [42–45]. The experiment was thus written as a pseudo 3D experiment, running five 2D experiments from the five pulse sequences found in Fig. 4.5. The linear combinations of the five resulting spectra that give rise to the S^3 HMBC subspectra are given by Eq. (4.1).

$$\underbrace{\underbrace{\left\{a \pm \frac{1}{2}[b+c]\right\}}_{\text{odd/even}} \pm \underbrace{\{d \pm e\}}_{\text{odd/even}}}_{S^3} \tag{4.1}$$

As indicated in Eq. (4.1) the terms $a \pm \frac{1}{2}[b+c]$ and $d \pm e$ are used to achieve multiplicity editing, while inter-brace addition/subtraction leads to S^3 editing of the active coupling constants. The active coupling is here defined comparable to DQF-COSY, where active couplings are anti-phase and passive in-phase, as the active coupling is observed between subspectra and the passive couplings are observed in a sub-spectrum, sometimes in the form of line-broadening. In the S^3 HMBC *homo* the active coupling is thus the long-range proton-proton coupling constants and the passive couplings are the long-range heteronuclear coupling constant and possibly coupling constants to other protons.

The spectra could in principle be combined in eight different manners but odd and even data are not mixed in practice, and the four are used where the sign in odd/even braces matches in Eq. (4.1). For clarification the spectrum with e.g. methine/methyl carbons from (a − ½[b + c] + [d − e]) will be referred to as "add" or "+" and (a − ½[b + c] − [d − e]) as "subtract" or "−" in this text.

The result is eight distinct HMBC spectra as given in Fig. 4.3, where four (5–8) are used for coupling constant extraction, and the other four may be used as standard multiplicity edited improved or HAT HMBC spectra. The sorting of the spectra is performed automatically by an in house developed AU-program which works seamlessly with Topspin, and the numbers in Fig. 4.3 correlate to spectrum numbering of the script.

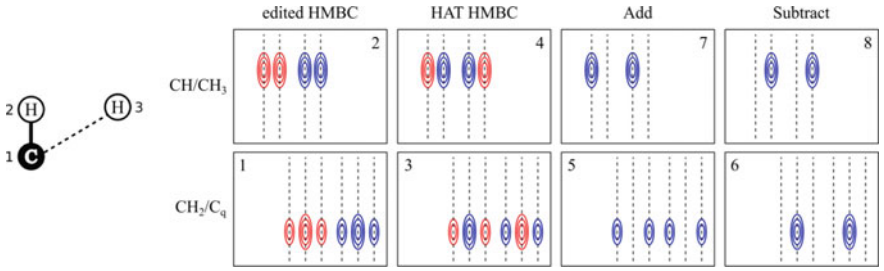


Fig. 4.3 The eight (numbered) output spectra from the S^3 HMBC *homo* experiment as given by the AU program. The chemical shift of carbon 1 is along F1, the chemical shift of proton 3 along F2 and the S^3 edited coupling constant of protons 2 and 3 is the displacement between the resonances in the add/subtract spectra. It is assumed that $J_{CH} > J_{HH}$ and that the multiplets shown are a CH···H(singlet) and a CH_2···H(singlet)

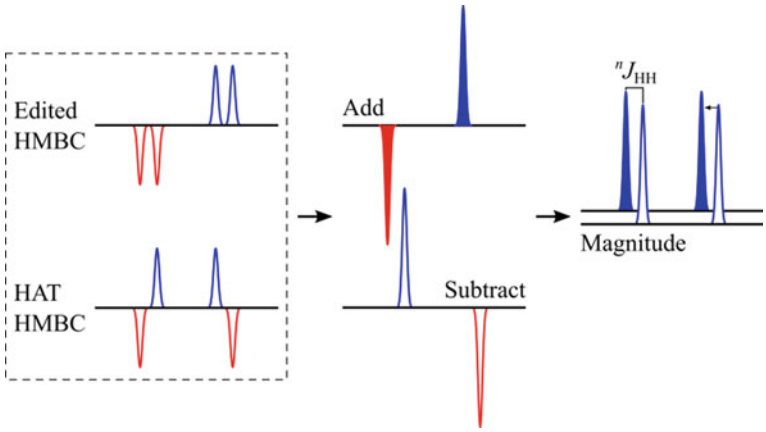


Fig. 4.4 Principle of the S^3 editing of CH–H in the S^3 HMBC experiment. Addition and subtraction of the edited-and HAT HMBC spectra leads to subspectra which, in magnitude mode, may be used to extract the $^{n}J_{HH}$-coupling constant. The second coupling constant present in the spectra is the $^{n-1}J_{CH}$-coupling constant. It is assumed that $J_{CH} > J_{HH}$ and that the multiplet is a CH···H(singlet)

The resulting subspectra are used to extract the long-range coupling constants including the value and relative sign. The coupling constants are determined from displacement (or direct measurements) of the peaks in matching +/− subspectra. The extracted coupling constant is between proton 2 in Fig. 4.3, attached to the carbon with chemical shift in F1, and the observed proton 3 with chemical shift in F2. The proton 3 and carbon 1 are coupled from the long-range heteronuclear coupling constant between the two, which is observed as a passive splitting in the spectra. A schematic representation of the S^3 editing is indicated in Fig. 4.3 and further explained in Fig. 4.4.

The absolute sign may be determined by keeping track of the S^3 add- and subtract spectra. The coupling is positive when the obtained subtract spectrum is positioned downfield relative to the add spectrum, if it is upfield the coupling constant is negative. This is governed by the sign of the $^1J_{CH}$-coupling constant which is always positive. For practicalities; if the subtract spectrum is used for displacement in Topspin, the size and sign of the coupling constants are extracted directly as presented in the program.

The five pulse sequences were used as previously published, with minor modifications to the HAT sequences [42–45]. In order to ensure equal length between the five sequences, delays were added. An initial delay of $\tau + \varepsilon'/2$ was added before, and $\tau/2$ just after, the first ^{1}H $\pi/2$ pulse of the HAT experiments, see the Fig. 4.5 caption for a definition of the delays. A decoupling period of $\tau + \varepsilon'/2 + {}^1$H $\pi/2$ was also implemented in accordance with the three standard HMBC experiments. Lastly, a delay of $\tau/2$ was added after the last ^{13}C $\pi/2$ pulse. This led to pulse sequences which uphold basic conditions for combinatory pulse sequences, as the time from first ^{1}H $\pi/2$ pulse to FID and the chemical shift and coupling constant evolution match between all five sequences. The magnetization thus experiences identical times of relaxation in all five sequences. The pulse sequences were programmed to run interleaved, with [b] and [c] spectra alternating to satisfy Eq. (4.1) without an initial need for scaling and thus an optimized run time versus *S*/*N*.

In practice the spectra are usually scaled though, due to high editing accuracy of the edited HMBC and HAT HMBC experiments. By employing the linear combinations [b + c] and [d ± e] the magnetization components in the sequences that could give rise to out-of-phase error terms are diminished. This results in slightly lower intensities for these spectra compared to (a), except for cases of CH_n pairs where τ in the sequences is exactly matched to $(2\ ^1J_{CH})^{-1}$. By scaling [b + c] by a constant k and [d ± e] by $(2k/[k + 1])^{½}$, the level of J cross talk is reduced. This is hardly needed due to an already low level of cross talk, but slight improvements were obtained when using a small scaling constant (usually in the order of $k = 1.05$). While it is possible to scale the cross-peaks individually to get optimal line shapes, this was not identified to be experimentally needed.

The add/subtract nature of the spectra is beneficial when considering the acquisition time, where four interleaved HMBC spectra may sound like a very long experiment to actually implement in normal acquisition setups. While a minimum of four scans are needed for each sequence, essentially setting a minimum time, the fact that the resonances are added to and subtracted from each other results in a gain of *S*/*N* in the final spectrum from every single scan. The broadband nature of the experiment and the fact that the coupling constants are measured in the F2 dimension also help to keep the acquisition time short, due to only having to run one experiment per sample and the inherent high resolution.

For the comparison of long-range coupling constants, especially those of very low absolute values, to theoretical values from 3D structures, a problem has often

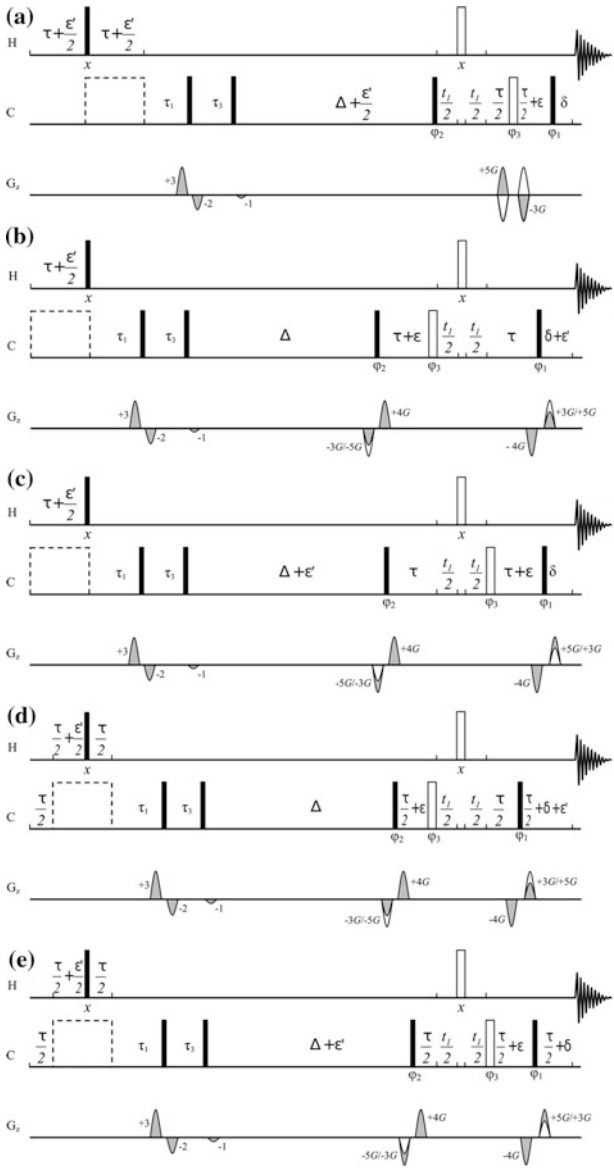


Fig. 4.5 S^3 HMBC *homo* pulse sequence comprised by **a**–**c**; the edited HMBC sequences, and **d**, **e**; the HAT HMBC sequences, all shown with a 2nd-order low-pass *J* filter (LPJF), which may be exchanged to 3rd-order LPJF or removed [1]. Reproduced with permission from the publisher [1]. Filled and open bars refer to π/2 and π pulses, respectively, and the dashed open boxes represent ^{13}C decoupling. $\tau = (2\ {}^1J_{CH})^{-1}$ or $({}^1J_{max} + {}^1J_{min})^{-1}$ δ is a gradient delay $\varepsilon = 2\ t_{1/2,min} + t(\pi_H)$ $\varepsilon' = \varepsilon + t(\pi_C)$ $\tau_1 = ½[{}^1J_{min} + 0.146\ ({}^1J_{max} - {}^1J_{min})]^{-1}$ $\tau_3 = ½[{}^1J_{max} - 0.146\ ({}^1J_{max} - {}^1J_{min})]^{-1}$ Δ is the delay for evolution under heteronuclear long-range couplings and is set to $(2 \cdot 8\ Hz)^{-1}$ as standard. Phase cycling is performed as $\varphi_1 = \{x, -x, -x, x\}$ $\varphi_2 = \{x, x, 4(-x), x, x\}$ $\varphi_3 = \{4(x), 4(y), 4(-x), 4(-y)\}$

been obtaining a high degree of accuracy. Thus dividing the coupling constants into categories such as low, medium, or high has been practiced [8]. This is in general a viable route for rigid structures, while it is problematic in e.g. conformational population determination. The accuracy of the extracted coupling constants is thus of utmost importance. In the following, extracted coupling constants will be compared to theoretical coupling constants with a high degree of correlation, even for very small coupling constants. To ensure that the displacements were not reliant on the human eye of the spectroscopist, in this case the author, a Matlab® script was developed which automatically identifies and extracts coupling constants from the spectra, see Appendix A6.

The pulse sequence was tested on multiple compounds, including strychnine and isopinocampheol (IPC), and selected results and points are included *vide infra*.

4.2.1 Computation of Long-Range Homonuclear Coupling Constants

The computation of homonuclear coupling constants may be achieved by two methods as previously discussed; Karplus type equations or density functional theory (DFT) methods [4, 5, 14]. No appropriate Karplus equations were identified for the long-range coupling constants and DFT methods were used. The taken approach is largely based on an article of Bally and Rablen, which establishes a good practice in calculating standard homonuclear coupling constants [14]. Since most studies of homonuclear coupling constants have focused on the easily accessible dihedral and germinal coupling constants, the applicability of methods and functionals to the long-range coupling constants needed to be evaluated, and S^3 HMBC *homo* data compared to multiple different functionals is found in Appendix A6.

In line with the results from Bally and Rablen, the optimal low-cost method of calculating homonuclear coupling constants was identified as optimization to B3LYP/6-31G(d), and calculation of the FC term of the coupling constants at a B3LYP/6-31G(d,p) level of theory. The function was augmented with four compact 1 s functions to more accurately calculate the FC term by using the input iop(3/10 = 1,100,000) in Gaussian [14]. The resulting *J*-coupling constants were scaled by 0.9117 in accordance with the literature [14]. The correlation between experimental and theoretical coupling constants is surprising in that the literature study included only few long-range coupling constants, and none as small as observed from S^3 HMBC *homo*—and still the fit is excellent.

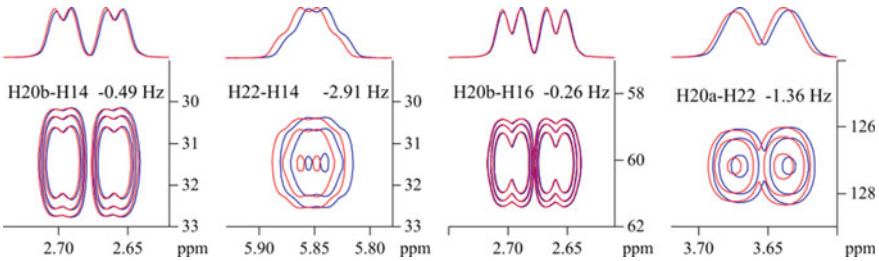


Fig. 4.6 Excerpts and cross-sections through F1 of cross-peaks of S^3 HMBC *homo* for strychnine (300 mM). Chosen as $n > 3$ for the associated $^nJ_{HH}$-coupling constants. Reproduced with permission from the publisher [1]

4.2.2 Results

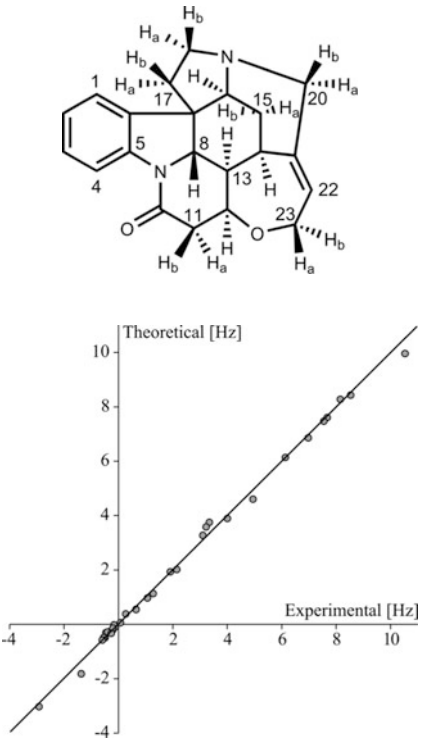
The S^3 HMBC *homo* experiment resulted in spectra from which the long-range coupling constants were easily extracted; see Fig. 4.6 for examples and a full spectrum in Appendix A6. Many $^3J_{HH}$-coupling constants were extracted, making the S^3 HMBC *homo* an excellent experiment for obtaining exact $^3J_{HH}$-coupling constants in spectra of compounds with a crowded proton dimension due to the extra resolution obtained by utilizing the carbon chemical shift range.

A total of 34 coupling constants were determined for strychnine. As a testament to the novelty of extracting small long-range homonuclear coupling constants, 13 coupling constants were extracted from this much utilized model compound, which were not reported in the literature. The extracted long-range coupling constants of strychnine are compared to theoretical values in Table 4.2.

It was only possible to obtain one coupling constant for some methine to methine correlations due to spectral overlap. When two coupling constants were identified, they were generally in good correlation with a maximum deviation of 0.22 Hz. In principle it is possible to extract coupling constants to methylene protons from a multiplet containing both coupling constants. In practice, the coupling constants to the methylene protons need to be large and different in order to determine coupling constants with a good correlation to the methine data, shown in Sect. 4.4.4. The general correlation of the data was excellent with a RMSD of 0.21 Hz for the full data set and, rather surprisingly, 0.15 Hz if only long-range coupling constants were considered. This was in part due to the values being lower, and relative to the size of the extracted coupling constants the differences were larger. It is also an indication of the high accuracy of the experiments though, and that the theoretical calculations that work well for $^3J_{HH}$ seemed to also correlate well to the long-range coupling constants. As indicated in Table 4.2, mostly 3- and 4-bond coupling constants were observed, and the novel coupling constants from the S^3 HMBC were all 4- or 5-bond and all below 1 Hz in magnitude.

Table 4.2 Left: Experimental and theoretical $^nJ_{HH}$-coupling constants of strychnine measured at 400 MHz ([a]800 MHz)

H1	H2	S^3 HMBC [Hz]	Theoretical [Hz]	n
1	2	7.68	7.61	3
1	3	1.50, 1.28	1.13	4
1	4	0.65, 0.66	0.55	5
2	4	1.07, 1.24	0.97	4
2	3	7.55[a]	7.47	3
3	4	8.16	8.28	3
8	**12**	**−0.16, −0.01**	**−0.10**	**4**
8	13	10.58, 10.54	9.96	3
8	**16**	**−0.32, −0.20**	**−0.27**	**4**
11a	12	8.54	8.43	3
11a	**13**	**−0.46**	**−0.31**	**4**
11b	12	3.35	3.75	3
12	13	3.23	3.58	3
12	**14**	**−0.20**	**−0.20**	**4**
12	**15a**	**−0.13**	**−0.09**	**5**
12	**23a**	**−0.10**	**−0.09**	**4**
12	**23b**	**−0.16**	**−0.02**	**4**
13	14	3.11	3.27	3
13	15a	0.27	0.38	4
13	**15b**	**−0.40**	**−0.28**	**4**
14	15a	4.95	4.60	3
14	15b	1.91	1.93	3
14	**16**	**0.65**	**0.54**	**4**
14	**20b**	**−0.49**	**−0.50**	**4**
14	22	−2.91	−3.03	4
15a	16	4.01	3.89	3
15b	16	2.15	2.02	3
16	17a	–	0.04	–
16	17b	–	−0.18	–
16	**17ab***	**−0.20**	**−0.14**	**4**
16	**18a**	**−0.54**	**−0.49**	**4**
16	**18b**	**0.08**	**0.06**	**4**
16	**20b**	**−0.26**	**−0.33**	**4**
20a	22	−1.36	−1.82	4
20b	22	−0.59	−0.58	4
22	23a	6.98	6.86	3
22	23b	6.14	6.14	3



Bold indicates coupling constants only published from S^3 HMBC. Theoretical coupling constants calculated by B3LYP/6-31g(d,p) u+1s from B3LYP/6-31g(d) optimized structures

*The protons of 17ab overlap and a combined value was used

Right Numbered structure of strychnine and experimental versus DFT coupling constants, where the line represents a perfect fit

Fit of data a = 1.01, b = 0.00, R^2 = 0.997

4.3 S^3 HMBC *Hetero*

An extended version of the homonuclear S^3 HMBC was developed to extract long-range heteronuclear coupling constants [2]. As already mentioned, the heteronuclear coupling constants are generally less utilized than their homonuclear counterparts, in part due to difficulty of extraction. The potential for structural information it huge however, as the number of $^{2-4}J_{CH}$-coupling constants will be larger than the number of $^{2-5}J_{HH}$-coupling constants for the majority of organic compounds. An example could be strychnine, a proton rich compound with the molecular formula $C_{21}H_{22}N_2O_2$, which has 96 possible $^{2-5}J_{HH}$ and 198 possible $^{2-4}J_{CH}$-coupling constants. In practice only a fraction of these coupling constants is measureable due to e.g. low coupling constants leading to no signal or overlap of chemical shifts. For proton deficient compounds, which are usually hard to solve, the larger abundance of heteronuclear coupling constants is more pronounced. In determining heteronuclear coupling constants, the sign-selective nature of the S^3 HMBC is imperative as $^2J_{CH}$ ranges from approximately −6 to 8 Hz while $^3J_{CH}$ ranges from 0 to 9 Hz [8].

4.3.1 Changes Needed

In order to change the observed coupling constant from $^nJ_{HH}$ to $^nJ_{CH}$ a new pulse sequence element needed to be implemented. This element needed to be able to exchange the polarization from proton to carbon (or from βα to αβ) in e.g. ^{13}C–^1H pairs, as illustrated in Fig. 4.7, where a zero-quantum pulse transfers the polarization from spin 2 (C–H) to spin 1 (C–H) to access the long-range heteronuclear coupling constant. A similar approach was used in the original S^3 work [46–49].

The new pulse train was established from the needed zero-quantum coherence polarization transfer. Theoretically there are two zero-quantum operators, $2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x$ and $2\hat{I}_y\hat{S}_x - 2\hat{I}_x\hat{S}_y$, and the operator $2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x$ was chosen [50, 51]. This needed to be translated into a pulse sequence, which was achieved by the following equations, starting from the ZQ_x operator in Eqs. (4.2) and (4.3). The transformations used are exemplified here, and explained further in Appendix A6.

$$ZQ_x = 2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x \tag{4.2}$$

$$e^{-i\pi ZQ_x} = e^{-i\pi\left(2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x\right)} = e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)}e^{-i\pi\left(2\hat{I}_x\hat{S}_x\right)} \tag{4.3}$$

The operator $e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)}$ was extended to Eq. (4.4), as $e^k e^{-k} = 1$.

$$e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)} = e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)}e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} \tag{4.4}$$

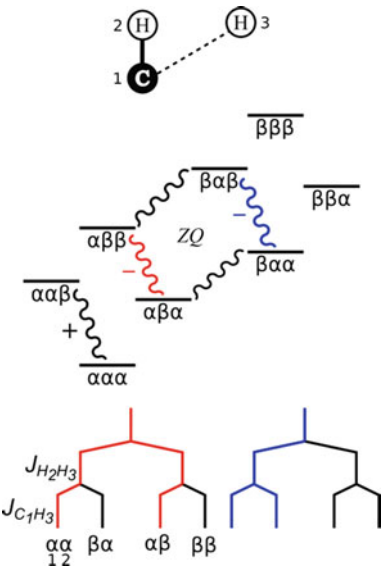


Fig. 4.7 Representation of the change in polarization between the two S^3 pulse sequences. Top: Relevant (numbered) spins. Middle: The energy levels (not to scale) of the 3 spins, the observed transitions in the homo (red) and hetero (blue) experiments (black transition is shared and used in both experiments) and the effect of the zero quantum (*ZQ*) pulse sequence. The associated linear combination is indicated by + and −. Bottom: The active couplings in the homo (red) and hetero (blue) experiment are indicated in a coupling tree for spin 3 for a case where $J_{HH} > J_{CH}$

The $e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)}e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}$ part was rewritten in Eq. (4.5) [51].

$$\begin{aligned} e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)}e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} &= e^{\left[-i\pi e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}\left(2\hat{I}_y\hat{S}_y\right)e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)}\right]} \\ &= e^{\left[-i\pi e^{i\frac{\pi}{2}\left(\hat{I}_x\right)}e^{i\frac{\pi}{2}\left(\hat{S}_x\right)}\left(2\hat{I}_y\hat{S}_y\right)e^{-i\frac{\pi}{2}\left(\hat{I}_x\right)}e^{-i\frac{\pi}{2}\left(\hat{S}_x\right)}\right]} \end{aligned} \tag{4.5}$$

Using that $e^{i\varphi B}Ae^{-i\varphi B} = A\cos\varphi - i[A, B]\sin\varphi$, the $e^{i\frac{\pi}{2}\left(\hat{S}_x\right)}$ $\left(2\hat{I}_y\hat{S}_y\right)$ $e^{-i\frac{\pi}{2}\left(\hat{I}_x\right)}$ part of Eq. (4.5) was equal to Eq. (4.6) [51].

$$e^{i\frac{\pi}{2}\left(\hat{I}_x\right)}\left(2\hat{I}_y\hat{S}_y\right)e^{-i\frac{\pi}{2}\left(\hat{I}_x\right)} = -i\left[\hat{I}_x, 2\hat{I}_y\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_y\right] = \left[2\hat{I}_z\hat{S}_y\right] \tag{4.6}$$

Similarly,

$$e^{i\frac{\pi}{2}\left(\hat{S}_x\right)}\left(2\hat{I}_z\hat{S}_y\right)e^{-i\frac{\pi}{2}\left(\hat{S}_x\right)} = -i\left[\hat{S}_x, 2\hat{I}_z\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_z\right] = \left[2\hat{I}_z\hat{S}_z\right] \tag{4.7}$$

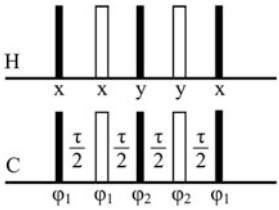


Fig. 4.8 The new zero quantum coherence pulse sequence element, which converts the polarization of the S^3 HMBC *homo* to achieve the *hetero* experiment. $\varphi_1 = \{x, -x, -x, x\}$, $\varphi_2 = \{y, -y, -y, y\}$

These answers were returned to Eq. (4.5), which was now equal to Eq. (4.8).

$$e^{\left[-i\pi e^{i\frac{\pi}{2}\left(\hat{I}_x\right)} e^{i\frac{\pi}{2}\left(\hat{S}_x\right)} \left(2\hat{I}_y\hat{S}_y\right) e^{-i\frac{\pi}{2}\left(\hat{I}_x\right)} e^{-i\frac{\pi}{2}\left(\hat{S}_x\right)}\right]} = e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} \tag{4.8}$$

Which led to the first operator in Eq. (4.3) being equal to Eq. (4.9).

$$e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)} = e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} \tag{4.9}$$

A similar approach was used throughout, done in Appendix A6, which led to Eq. (4.10).

$$\begin{aligned} e^{-i\pi\left(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x\right)} &= e^{-i\pi\left(2\hat{I}_y\hat{S}_y\right)} e^{-i\pi\left(2\hat{I}_x\hat{S}_x\right)} \\ &= e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} e^{-i\frac{\pi}{2}\left(\hat{I}_y+\hat{S}_y\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{i\frac{\pi}{2}\left(\hat{I}_y+\hat{S}_y\right)} \\ &= e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{-i\frac{\pi}{2}\left(\hat{I}_y+\hat{S}_y\right)} e^{i\frac{\pi}{2}\left(\hat{I}_z+\hat{S}_z\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{i\frac{\pi}{2}\left(\hat{I}_y+\hat{S}_y\right)} \\ &= e^{-i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{-i\frac{\pi}{2}\left(\hat{I}_y+\hat{S}_y\right)} e^{-i\pi\left(2\hat{I}_z\hat{S}_z\right)} e^{i\frac{\pi}{2}\left(\hat{I}_x+\hat{S}_x\right)} \end{aligned} \tag{4.10}$$

Finally, it was used that –x = 270° x = 90° x + 180° x etc. to convert the operators into the pulse train seen in Fig. 4.8. The optimal delays between the new pulses are $\tau/2$ or $(4\ {}^1J_{CH})^{-1}$, in accordance with polarization transfer of methine groups [50].

4.3.2 Pulse Sequences

The initial proposal for the S^3 HMBC *hetero* was to substitute the final ^{13}C $\pi/2$ pulse in the original S^3 HMBC *homo* experiment with a new pulse train, as seen in Fig. 4.9. Note that the first $\pi/2$ pulse is omitted in the carbon channel, since the carbon spins need to be aligned along z before detection.

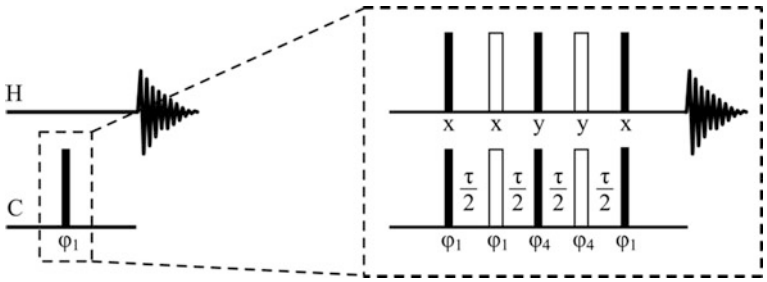


Fig. 4.9 The extended pulse sequence which replaces the final π/2 S_x pulse of the S^3 HMBC to get the *hetero* experiment. $\varphi_1 = \{x, -x, -x, x\}$, $\varphi_4 = \{y, -y, -y, y\}$

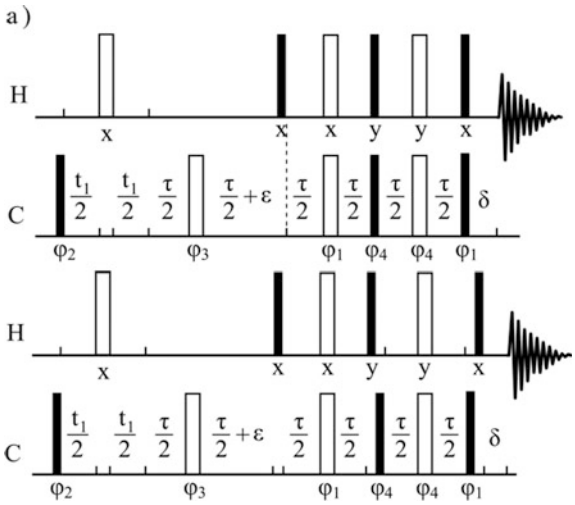


Fig. 4.10 Representation of two possible methods to obtain the needed adjustments of the delays for carbon chemical shift refocusing, exemplified for sequence (**a**)

In addition to inserting the new pulse sequence, the gradients of the S^3 HMBC *homo* experiment (a) in Fig. 4.5, had to be adjusted, by extension to select $G_1(H^{-1}/C^{+1})$–$G_2(H^{-1}/C^{-1})$–$G_3(H^{-1}/C^{0})$ instead of $G_1(H^{-1}/C^{+1})$–$G_2(H^{-1}/C^{-1})$. This was achieved by using a gradient power ratio of 2:−2:1 instead of 5:−3 for echo, and −2:2:1 compared to −3:5 for anti-echo. Upon appending the new pulse train to obtain a new pulse sequence it was necessary to ensure that the carbon chemical shifts are refocused between the first ^{13}C π/2 (disregarding the low pass *J*-filter), and the first ^{13}C π/2 pulse in the new pulse train. This was in principle achieved in two ways which should both lead to refocusing, as illustrated in Fig. 4.10, using sequence (a) as an example. In short, the first version withdraws the time of a ^{1}H π/2 pulse from the delay after the first ^{13}C π pulse to ensure refocusing of δ_C, while the π/2 pulses in the pulse train are offset in the second version in a manner that ensures refocusing.

General rules when combining data for pulse sequences needed to be obeyed as well:

1. The ^{13}C chemical shift was refocused for the minimum t_1.
2. Sample heating due to decoupling was equal.
3. The evolution times for δ_H, J_{HH} and J_{CH} were equal.
4. The lengths of the sequences were identical.

All of these requirements were met in the S^3 HMBC, and since an identical pulse train was appended, all should be met by design.

Unfortunately this strategy was not successful. While the polarization transfer worked and led to extraction of $^nJ_{CH}$-coupling constants, the refocusing of carbon chemical shifts was imperfect. The refocusing problems for methine groups were negligible and resulted only in slight broadening of the resonances in the F1 dimension in the add/subtract spectra. For quaternary carbons the problem was more pronounced, which is exemplified in Fig. 4.11, and even though these are of no interest in this experiment, the problem needed to be solved, to ensure the extraction of correct coupling constants throughout. Note also that the elimination of quaternary carbons in the methine/methyl spectra by the multiplicity editing did not work properly as a consequence.

The next approach taken was to append the new pulse train after the entirety of the original homonuclear S^3 HMBC, as the refocusing here worked perfectly, leading to inherent refocusing of the new pulse sequence. The gradients were rearranged to achieve an assumed clean coherence selection as seen in Fig. 4.12, top left. This adjustment did not initially give the desired refocusing of carbon chemical shifts though. The gradient selection was identified to be the cause of the problem, as the top right sequence in Fig. 4.12 led to messy but refocused spectra.

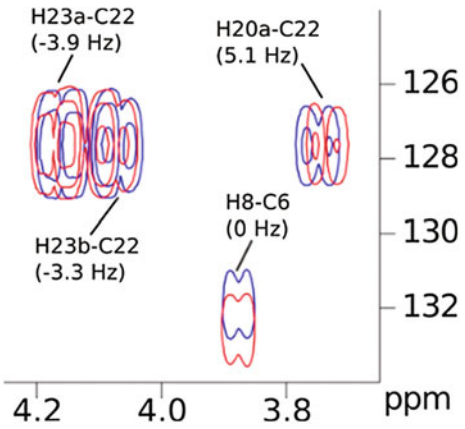


Fig. 4.11 Example of methine and quaternary carbon displacement in the S^3 HMBC *hetero* methine/methyl spectrum of strychnine (180 mM)

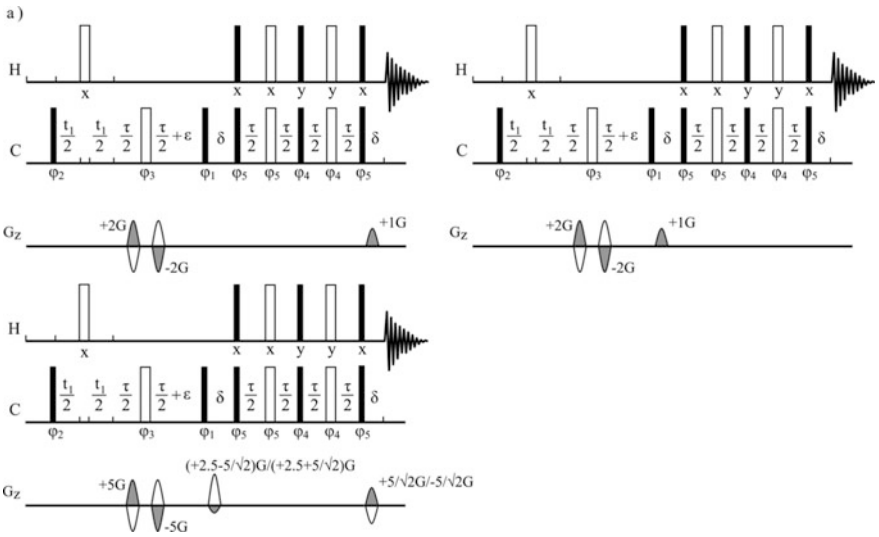


Fig. 4.12 Appending the pulse train to the S^3 HMBC *hetero* sequences. Sequence **a** used for example. Top left: Initial attempt of appending the new sequence which leads to displacement of ^{13}C. Top right: Second attempt, with no displacement of ^{13}C but more noise (expected). Bottom left: Final sequence; no displacement and *S/N* in level of S^3 HMBC *homo*

Thus, a new gradient scheme was developed in order to alleviate this problem, as illustrated in Fig. 4.12 bottom left for sequence (a), while similar corrections were made for sequence (b–e). Note that a similar gradient scheme was not possible for the first proposed pulse sequence, due to a mismatch of gradient duration and delays in some of the sequences.

It is not clear why the new gradient scheme is superior, as the chosen coherence pathway is supposedly identical between all sequences. The coherence selection by the gradients in the two first sequences in Fig. 4.12 is $G_1(H^{-1}/C^{+1})$–$G_2(H^{-1}/C^{-1})$–$G_3(H^{-1}/C^{0})$. In the final sequence an extra gradient pair was added to choose coherence $G_1(H^{-1}/C^{+1})$–$G_2(H^{-1}/C^{-1})$–$G_3(H^{-1}/C^{0})$–$G_4(H^{-1}/C^{0})$, or in words, make sure that the change in coherence order was zero for both protons and carbons across the new pulse train. Experimental evidence showed that it was needed to gain this further control of the coherence selection, which in practice was achieved by using gradients as in the S^3 HMBC, only changing sequence (a) as earlier described, and inserting new gradients around the new pulses with opposite signs. The first new gradient is added to the last existing gradient from the S^3 HMBC *homo* sequence, as indicated in Fig. 4.12.

Appending the new pulse train led to an increase in the phase cycle compared to just exchanging the last ^{13}C $\pi/2$ pulse, as the phases of the new element were cycled independently of φ_1. This was not strictly needed as the spectra and coupling constants using $\varphi_1 = \varphi_5$ and $\varphi_1 \neq \varphi_5$ were essentially identical. It is thus

recommended that for four scan (ns) experiments φ_1 is set equal to φ_5, while φ_1 should be cycled independently of φ_5 for ns > 4 scans. In both cases φ_2 and φ_3 were cycled independently and should be changed accordingly. Since the pulse train was appended and identical for all pulse sequences, previously mentioned requirements for combining pulse sequences were inherently met.

The new pulse sequences are found in Fig. 4.13 and the result of the change in polarization is exemplified in Fig. 4.14 by the simple compound vinyl acetate. The coupling constants are compared to 1D and reference values to an excellent correlation in Fig. 4.15 and Table 4.3 [52].

To better understand the coupling pattern of the homo- and heteronuclear variants of the S^3 HMBC the cross-peaks of C-3/H-4b and C-3/H-4a are explained in Fig. 4.15. As seen both J_{CH} and J_{HH} were part of the multiplet in both experiments and if the (+) and (−) spectra were overlaid, the combined shapes of the multiplets were almost identical.

The in-phase J_{H4bH4a} coupling constant was determined to be larger for both experiments than the 1.5 Hz determined from a ^{1}H spectrum, possibly due to a lack of resolution of the S^3 HMBCs with 2.5 Hz/point (4 k vs. 64 k points acquired for 1D ^{1}H spectrum). As expected the coupling trees were alike with the difference being the coupling constant leading to S^3 editing, where sign determination was possible.

J-crosstalk was experimentally more apparent, for a few resonances, compared to the native S^3 HMBC where *J*-crosstalk was always in the noise level. The *J*-crosstalk may alter the experimentally determined coupling constants and will always lead to the extraction of smaller coupling constants and needed to be addressed. As for the homonuclear experiment, the *J*-crosstalk could be minimized by scaling [b + c] with a factor *k* and [d + e] with $(2\ k/[k + 1])^{½}$. If scaling was not sufficient the crosstalk was diminished by scaled combinations of the α and β spectra, as described earlier for S^3 experiments [41]. The subspectra of vinyl acetate were the only spectra where *J*-crosstalk was immediately apparent, even when subspectra of other compounds were scrutinized. In Fig. 4.16 the standard editing of the spectra is compared to spectra edited by a scaled combinations of the α and β spectra to decrease *J*-crosstalk [41]. Most importantly the extracted *J*-coupling constants varied by less than 0.1 Hz, and thus the differences were lower than the expected experimental uncertainty of these kinds of experiments. It is acknowledged that the difference may be larger for different compounds as described in the literature [53].

4.3.3 Computation of Long-Range Heteronuclear Coupling Constants

Unlike the homonuclear long-range coupling constants, the heteronuclear variants have been more extensively investigated [8, 9, 21, 22, 54]. Unfortunately, a study

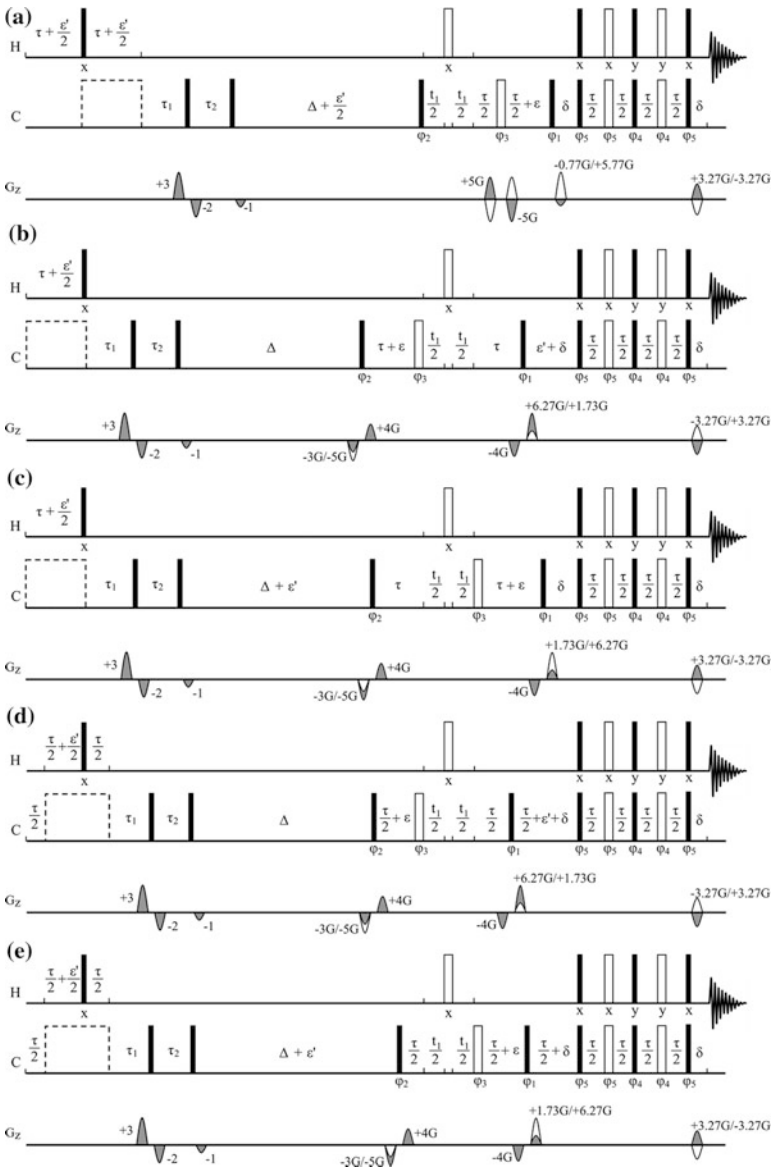


Fig. 4.13 Next page. S^3 HMBC *hetero* pulse sequence comprised by modified sequences **a–c**; edited HMBC sequences, and **d**, **e**; HAT HMBC sequences, all shown with a 2nd-order LPJF, which may be exchanged to 3rd-order LPJF or removed. Reproduced with permission from the publisher [2]. Filled and open bars refer to π/2 and π pulses, respectively, and the dashed open boxes represent ^{13}C decoupling. $\tau = (2\ ^1J_{CH})^{-1}$ or $(^1J_{max} + {}^1J_{min})^{-1}$, δ = gradient delay, $\varepsilon = 2\ t_{1/2,min} + t(\pi_H)$, $\varepsilon' = \varepsilon + t(\pi_C)$, $\tau_1 = ½[^1J_{min} + 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$, $\tau_2 = ½[^1J_{max} - 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$, Δ = delay for heteronuclear long-range coupling evolution. $(2 \cdot 8\ Hz)^{-1}$ is standard. Phase cycles. $\varphi_1 = \{x, x, 4(-x), x, x\}$, $\varphi_2 = \{4(x), 8(-x), 4(x)\}$, $\varphi_3 = \{8(x), 8(y), 8(-x), 8(-y)\}$, $\varphi_4 = \{x, -x, -x, x\}$, $\varphi_5 = \{y, -y, -y, y\}$

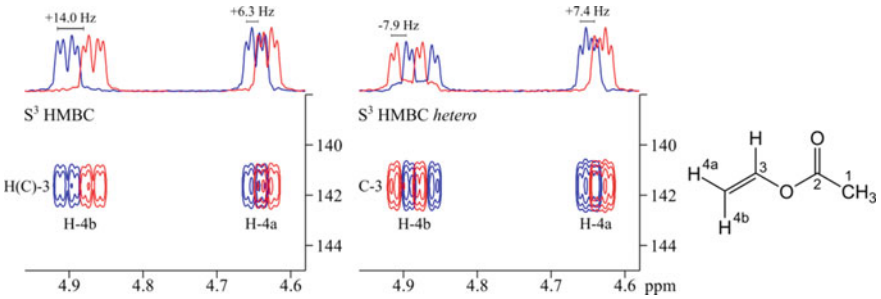


Fig. 4.14 Left: Comparison of S^3 HMBC *homo* and S^3 HMBC *hetero* spectra for C(H)-3 of vinyl acetate (220 mM) Reproduced with permission from the publisher [2]

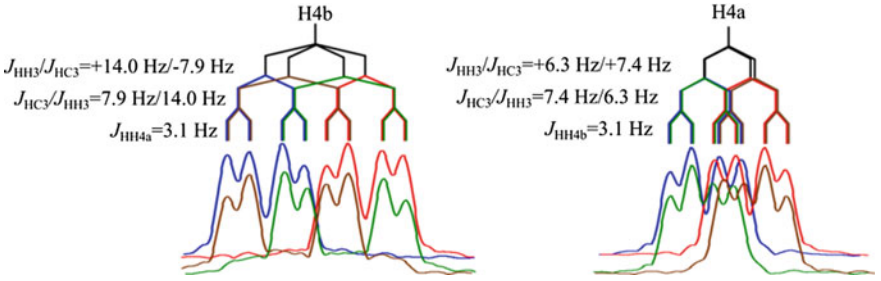


Fig. 4.15 Comparison of S^3 HMBC *homo* (blue/red) and S^3 HMBC *hetero* (green/brown) multiplet for H-4b and H-4a of vinyl acetate. Black lines in coupling tree indicate S^3 editing. The coupling trees are offset slightly horizontally to better distinguish colors

Table 4.3 Comparison of S^3 HMBC coupling constants of vinyl acetate to 1D ^{1}H spectra (*a*) and literature (*b*) [52]

	$^3J^a_{HH}$ [Hz]		$^2J^b_{CH}$ [Hz]	
	S^3 HMBC *homo*	Ref.	S^3 HMBC *hetero*	Ref.
H-4a	+6.3	6.2	+7.4	+7.6
H-4b	+14.0	14.0	−7.9	−7.9

similar to that of Bally and Rablen was not found, and the reported calculations use a wide variety of methods and functionals, some of which seem to benefit from scaling [14, 54]. This is in principle not a problem, but the lack of a true and tried methodology and scaling factor may lead to overfitting as the computed coupling constants are scaled to fit only the current data or a couple of datasets. However the method of "local scaling" is generally used in the calculation of chemical shifts, as discussed in Sect. 2.4.3 [55–57].

Investigating and correlating calculations of long-range heteronuclear coupling constants to a large set of experimental data was not a part of the work; in part due

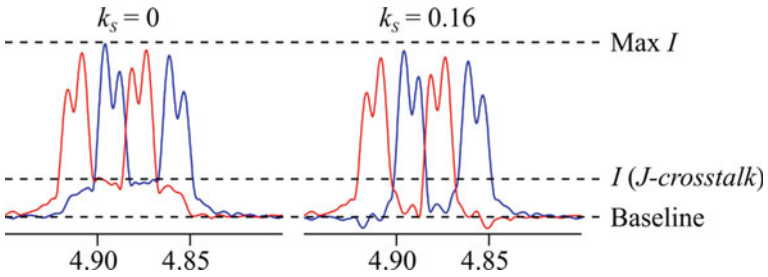


Fig. 4.16 Comparison of the *J*-crosstalk in standard and edited spectra of the C3-H4a cross-peak of vinyl acetate (220 mM). Edited by $\alpha_S = \alpha - k_S \cdot \beta$ and $\beta_S = \beta - k_S \cdot \alpha$. The difference in the extracted J_{CH}-coupling constant is below 0.1 Hz

to discrepancies between methods to extract coupling constants. It should be noted though, that scaling is often required for accurate determination of NMR observable from DFT calculation, probably due to a constant and inherent error in the calculations. It is thus often advantageous to use the correlation factor to assess the fit between experimental and theoretical data, as the correlation factor is independent of needed scaling.

Two methods were investigated with different functionals; B3LYP, which worked nicely for homonuclear coupling constants, and MPW1PW91, which is very often used specifically to better calculate NMR properties [8, 14, 55, 57–59]. For both methods it was apparent that a larger basis set improved the correlation between experimental and calculated data, while e.g. diffuse functions had little to no effect. It was necessary to use the full theoretical *J*-coupling constants and not just the Fermi contact (FC) to get the best results, while the FC term was still dominating in the calculations. This was in contrast to the method used for calculations of J_{HH} by DFT. For more on different components used in *J*-coupling constants calculations by DFT see Sect. 2.4.4.

MPW1PW91 generally resulted in a better correlation to the experimental data when used to calculate coupling constants while the difference between optimized structures from B3LYP and MPW1PW91 differed little, as long as identical basis sets were used. In the end MPW1PW91/6-311G(d,p) was used for both optimization and *J*-coupling constant calculations as this resulted in the better reproduction of the experimental *J*-coupling constants. Comparisons may be found in Appendix A6.

4.3.4 Results

The S^3 HMBC *hetero* experiment was as the homonuclear variant tested on strychnine. The coupling constants were easily extracted from displacements of 1D

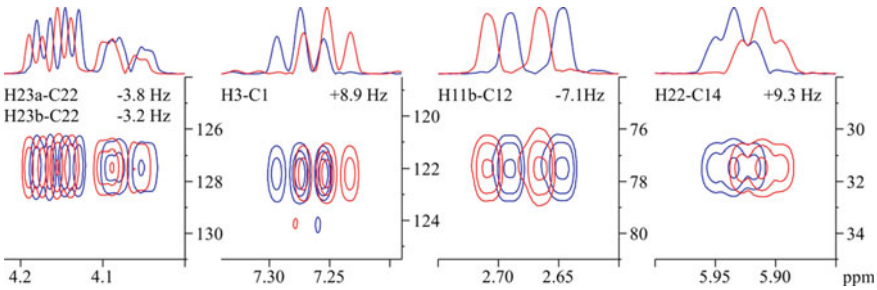


Fig. 4.17 Excerpts and cross-sections through F1 of cross-peaks of S3 HMBC *hetero* for strychnine (180 mM). Note that the sign of the coupling constants are apparent from the relative position of the cross-peaks in the subspectra Reproduced with permission from the publisher [2]

slices from cross-peaks equivalent to the homonuclear S^3 HMBC. Examples of cross-peaks as well as 1D slices are found in Fig. 4.17 and a full spectrum in Appendix A6.

The extracted coupling constants correlate well to theoretical coupling constants, especially if the calculations are scaled, and a comparison to theoretical as well as literature values is presented in Table 4.4 and Fig. 4.18.

The extracted coupling constants were in excellent agreement with previously published data and scaled theoretical calculations, with an RMSD of 0.22 Hz when comparing to the latter. This is almost equal to the RMSD determined for homonuclear coupling constants from S^3 HMBC (0.21 Hz), indicating the same general precision of the two methods. When comparing to the theoretical calculations, note also the high correlation factor. This is in support of the differences between experimental and calculated coupling constants being due to an inherent and consequent error in the DFT calculations, which may be alleviated by scaling consistent with literature findings [14]. A conclusion, which was supported by differences in scaling factors a and b between functionals, while a high R^2 was generally retained. It could of course also be the product of a constantly erroneous extraction of coupling constants but good correlation to literature values contradicts this [29, 36, 60, 61]. It would have been beneficial to have a golden standard for comparison as to determine whether the deviation was a result of the experiment or the theoretical method utilized. Like the homonuclear S^3 HMBC it was possible to extract coupling constants across heteroatoms, which is important in e.g. carbohydrate 3D structural investigations [9, 62, 63]. This is a nice feature of HMBC compared to TOCSY type experiments, as the latter utilizes $^3J_{HH}$ to generate the long-range coupling constants [21].

The S^3 HMBC *hetero* experiment, and in large regard also the homonuclear S^3 HMBC experiment, has one inherent flaw; it is only really a viable option for coupling constants involving methine carbons. While this may limit widespread

Table 4.4 Experimental S^3 HMBC *hetero*, literature and theoretical $^nJ_{CH}$-coupling constants for strychnine [29, 36, 60, 61]

H	C	Experimental (Hz)					Theoretical (Hz)		n
#	#	S3 HMBC *hetero*	[36]	[60]	[29]	[61]	Calc.	Scaled	
1	3	7.4	7.5, 7.2				6.7	7.4	3
2	4	7.9					7.0	7.8	3
3	1	8.9	8.9, 9.3				8.1	8.9	3
4	2	7.7		7.5,7.4	7.9		7.1	7.8	3
8	12	6.4	5.8, 5.6	5.5, 5.6	5.3		5.6	6.2	3
8	13	−1.7					−1.7	−1.6	2
11b	12	−7.1	6.9, −6.9	7, 6.9	6.8		−6.9	−7.3	2
13	8	−6.3		6.3, 6.4, 6.3	6.2		−6.0	–6.3	2
13	14	−4.9	5.5, −4.5	4.6, 4.7	4.5	4.7	−4.7	−5.0	2
15a	13	7.7		8,8.1	8.0	7.9	7.0	7.7	3
15a	14	−2.9	1.2, −2.9	1.8, 2.8, 2.8	2.0	3.2	−2.5	−2.6	2
15a	16	−4.6		4.5, 4.3	4.0		−4.3	−4.5	2
15b	14	−3.2	2.3, −2.6	3.2, 3.6	2.8	4.8	−3.1	−3.2	2
15b	16	−1.9		2.7, 3	1.4		−2.1	−2.2	2
16	14	6.2	6.7, 6.4			6.4	5.8	6.4	3
20a	14	1.4					1.3	1.6	3
20a	22	5.6	4.7, 5	6.1, 5.8	5.1		4.6	5.1	3
20b	14	5.5	5.7, 5.4	5.4, 5.3	5.1	5.5	4.7	5.2	3
20b	16	7.8		6.9, 6.9	6.7		6.3	6.9	3
20b	22	4.6	5.1, 5.5	4.5, 4.6	4.2		4.6	5.1	3
22	14	8.5	8.5, 8.6	7.9, 8.5, 8.6	8.8	8.9	8.0	8.8	3
23a	22	−3.8	4.2, −3.8	3.8, 4			−3.8	−3.9	2
23b	22	−3.4	3.6, −3.4	4, 3.9			−3.0	−3.1	2
					RMSD		0.53	0.22	
					R^2		0.999		

All *J*-coupling constants are in Hz. Both optimization and *J*-coupling constant calculations were performed at a MPW1PW91/6-311+g(d,p) level of theory. Calculated coupling constants are given without scaling and linearly scaled to the experimental data as $J_{scaled} = (J_{calc} - b)/a$. $a = 0.92$, $b = -0.14$. Reproduced with permission from the publisher [2]

usage, the high accuracy, ease of extraction and sign-selective nature of the experiment should cater for a lot of fields where the resulting information will immediately increase the structural knowledge. An obvious field is that of carbohydrate chemistry, where coupling constants across the glycosidic linkage may be used for conformer analysis [9, 62, 64, 65].

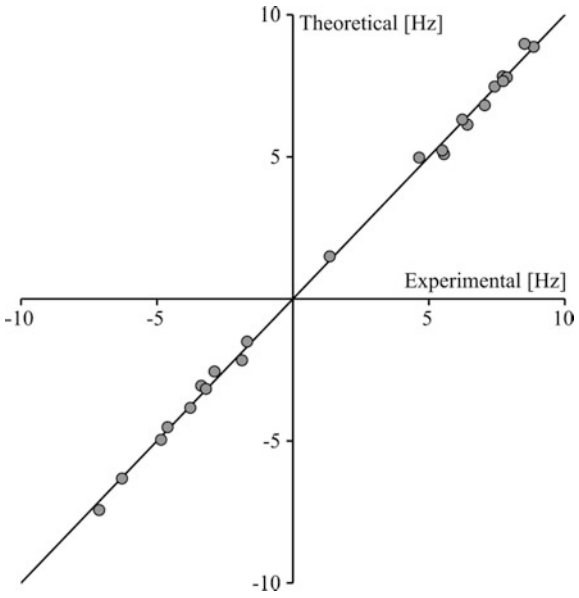


Fig. 4.18 Experimental versus DFT-calculated coupling constants, where the line represents a perfect fit. Calculated coupling constants are linearly scaled to the experimental data as $J_{scaled} = (J_{calc} - b)/a$. $a = 0.92$, $b = -0.14$. Reproduced with permission from the publisher [2]

4.4 Perspectives

As interesting as the possibility to obtain the long-range coupling constants was, the greater question is whether the observed long-range coupling constants could actually discern stereoisomers. The determined long-range J_{HH} values were small and the information may be non-discerning and thus of lesser value. To assess this, a probability function was considered attractive.

4.4.1 Establishing a Probability Function for J

The DP4 is in the opinion of the author, a very informative way to evaluate diastereomers from chemical shift differences, see Sect. 3.1.5 [66]. A similar probability function, here dubbed JP4, would be desirable for *J*-coupling constants especially as DFT calculations become easier available.

A probability function was thus established from coupling constants obtained from the literature, as determining these from scratch was not in the scope of the project and was deemed too time consuming. The dataset published by Bally and Rablen was used as the combined dataset of test and the two probe sets (combined to 228 coupling constants from multiple structures) [14]. The standard deviation from the *J*-coupling constants of their dataset is 0.51 Hz [14]. The intercept of the

dataset was 0.0026 Hz, and is approximated to 0 Hz in the following. The degrees of freedom (v) parameter was harder to determine and it was only estimated in the current work.[1] The data set was fitted to a standard student t-test curve to a value where the difference in the fit of v and $v - 1$ was below 0.01. This equals a value of 8.76 ($\sim$9). It should be noted that v does not drastically influence the result of the evaluation, except when a very low or extremely high value is chosen. The formula for the resulting "JP4" probability function is given in (4.11).

$$P(i|J_N) = \frac{\prod_{k=1}^{N} 1 - T^v\left(\left|J_{scaled,k}^{-i} - J_{exp,k} - \mu\right|/\sigma\right)}{\sum_{j=1}^{m}\prod_{k=1}^{N} 1 - T^v\left(\left|J_{scaled,k}^{-i} - J_{exp,k} - \mu\right|/\sigma\right)} \tag{4.11}$$

With $\mu = 0$, $v = 9$ and $\sigma = 0.51$ Hz. The number of data points compared to the DP4 probability function was rather sparse with 228 J-coupling constants compared to 1717 $\delta(^{13}C)$ and 1794 $\delta(^{1}H)$, and thus the resulting probability will be prone to more errors. The JP4 value will thus never be used alone, but it is a very intuitive way to compare experimental and theoretical data sets. It should also be noted that this is of course only valid if the methodology of Bally and Rablen is used [14]. The output of the JP4 function is the probability in percent that a given stereoisomer is correct according to the data.

4.4.2 *Differentiation of Stereoisomers by S³ HMBC*

To check whether the long-range coupling constants were discerning of stereoisomers the data was purposely calculated for wrong assignment of diastereotopic protons for strychnine.

Starting from the homonuclear coupling constants the correlation between the full and a reduced set of experimental coupling constants and theoretical values are seen in Table 4.5. The reduced set is the data with all $^{3}J_{HH}$-coupling constants removed, as these could on the most part be extracted with other methods.

The mean average error (MAE) was not necessarily a good method to differentiate the datasets as it was hard to evaluate whether the differences were within expected errors. Here the JP4 function made it much easier to quickly evaluate the data. Even without the $^{3}J_{HH}$-coupling constants it was possible to assign all diastereotopic proton pairs using the JP4 probability function except for two: 17, where the protons overlap, and 23, where the two observed coupling constants available were almost equal (both expected to be positive and under 0.1 Hz). It should be noted that, as it is a probability function, the values of the right assignment and the misassignment of 23 meant that they could not be discerned.

[1] v was determined from a statistics program based on the data set in the original DP4 publication, which was not available to the author. $v(^{1}H) = 11.38$, $v(^{13}C) = 14.18$ [66].

Table 4.5 Comparison of correct and incorrect assignment of the diastereotopic protons in strychnine

	Full		Reduced	
Switched #	MAE	JP4	MAE	JP4
Right	0.15	99.9	0.06	47.9
Dia-11	0.38	0.0	0.08	0.9
Dia-15	0.38	0.0	0.09	0.0
Dia-18	0.17	0.1	0.09	0.0
Dia-20	0.23	0.0	0.15	0.0
Dia-23	0.18	0.0	0.06	51.2
Dia-All	0.75	0.0	0.22	0.0

The full set includes all coupling constants extracted by S^3 HMBC *homo*, while $^3J_{HH}$-coupling constants have been removed in the reduced set. The number on the left indicates the diastereotopic protons which are switched compared to the correct assignment

Table 4.6 Comparison of right and wrong assignment of the diastereotopic protons in strychnine using $^nJ_{CH}$-coupling constants from the S^3 HMBC *hetero* experiment

#	MAE	JP4
–	0.24	94.1
11	0.88	0.0
15	0.66	0.0
20	0.75	0.0
18	1.02	0.0
23	0.29	5.9
All	2.63	0.0

Still, it was a testimony to the high degree of accuracy and the importance of identifying the sign in long-range coupling constant experiment that the remaining diastereomers could be differentiated, underlining the vast possibilities for using this experiment for 3D structural elucidation. All of the coupling constants in the reduced set were thus between −3.0 and 1.1 Hz, and very exact coupling constants were needed to facilitate differentiation.

In a similar fashion the diastereotopic protons of strychnine were again switched for the heteronuclear data, and the experiment was clearly discriminating, as seen in Table 4.6.

It is recognized that using the JP4 probability function for comparison is clearly problematic; and that the value holds little meaning. But, as the standard deviation from the strychnine data was smaller than for the dataset used in the setup of the JP4 probability (and equal to coupling constants from the homonuclear S^3 HMBC variant) [14], and if a comparable degrees of freedom value is assumed, it is indicative of the discriminative prowess of the long-range J_{CH}-coupling constants.

The $^nJ_{CH}$- and $^{n+1}J_{HH}$-coupling constants were clearly discriminative of diastereomers and the two experiments are believed to greatly increase the amount of available structural data and resulting structural knowledge.

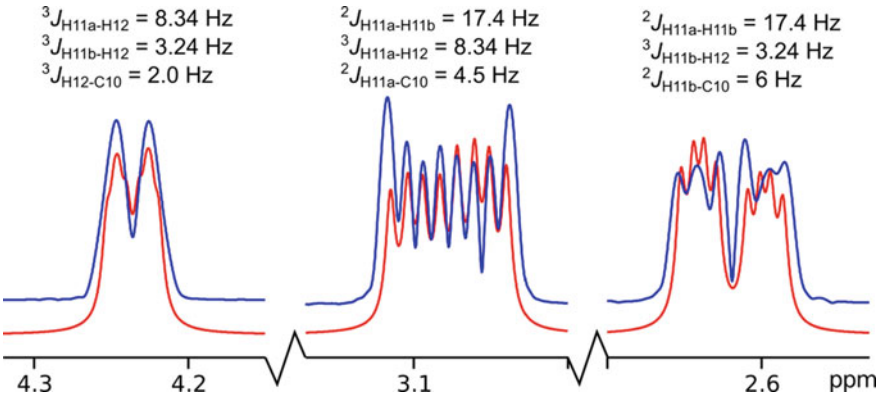


Fig. 4.19 1D slices of couplings to C-10 in strychnine (300 mM) from S^3 HMBC *homo*. Experimental data (blue) compared to simulated spectra (red). Theoretical heteronuclear coupling constants are (from left): 1.2, −5.1 and −7.3 Hz. Spectral resolution in the S^3 HMBC *homo* was 2.5 Hz/point. Simulated using the coupling constants given above the resonances with a line broadening of 2.5 Hz using the Daisy utility in Topspin

4.4.3 $^nJ_{CqH}$-Coupling Constants from S^3 HMBC Homo

The major drawback of the S^3 HMBC *hetero* experiment was the inability to measure coupling constants to methylene or quaternary carbons. Especially the latter is problematic, as the structural knowledge gained from $^nJ_{CH}$-coupling constants is increasingly needed as the proton-to-heteroatom ratio decreases. In that regard it is appreciable that J_{CqH} coupling constants are implicitly embedded in the add spectrum of methylene and quaternary carbons in both S^3 HMBC experiments, without sign due to the lack of S^3 editing.[2] The multiplet became complicated in cases where the proton couples to other protons, but in proton deficient compounds where the proton multiplets should be simple, maybe even singlets, the size of the coupling constants may be easily extracted and used in structure elucidation. An example of a rather complicated situation is given in Fig. 4.19 where 1D slices from S^3 HMBC *homo* of strychnine are compared to theoretically simulated spectra using the given coupling constants. Theoretical heteronuclear coupling constants are, in order from left: 1.2, −5.1 and −7.3 Hz, calculated at a MPW1PW91/6-311+g(d,p) level of theory and scaled according to the methine data. The extracted coupling constants could thus be used in a large-medium-small analysis as suggested by Bifulco et al. if not used directly [8]. In-phase homonuclear coupling constants were initially those from Table 4.2 and varied slightly to increase the fit between the experimental and theoretical spectra. The $^2J_{HH}$-coupling constant was estimated from 1D ^{1}H spectra.

[2]Quaternary carbons cannot have their polarization interchanged from ^{1}H to ^{13}C, for obvious reasons.

4.4.4 Extraction of Coupling Constants to Methylene (C)H

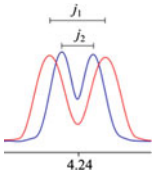
Unlike the easy extraction of coupling constants to methine groups, extraction of constants to methylene and methyl groups is more problematic, making this experiment a poorer choice for extractions of this kind of data. It seldom posed a significant problem in practice, since coupling constants that involve a methine and e.g. a methylene group in most cases could be extracted easily from the methine carbon. A simple approach to extract approximate coupling constants to methylene protons was identified, and given in Eq. (4.12), from measured distances between in-phase peaks (j_1 and j_2), illustrated in Table 4.7. It is especially easy when the passive coupling constants were small.

$$ {}^{n}J^{1}_{HH} = \frac{j_1 + j_2}{2}, \quad {}^{n}J^{2}_{HH} = \left| {}^{n}J^{1}_{HH} - j_2 \right| \tag{4.12} $$

The equation yields the coupling constants as two active coupling constants lead to S^3 editing per coupling pattern, while the remaining coupling constants are in-phase and thus appear in both subspectra. The overlaid cross-peaks will thus have a pattern where the splitting in one sub-spectrum is equal to ${}^{n}J^{1}_{HH} + {}^{n}J^{2}_{HH}$ and the other $\left| {}^{n}J^{1}_{HH} - {}^{n}J^{2}_{HH} \right|$. Coupling constants extracted by this method are given in Table 4.7 and compared to methine data when possible.

In Fig. 4.20 examples of simulated spectra based in coupling constants to methylene protons are depicted, from the determined values found in Table 4.7. When the coupling constants were large and different they were easily extracted, as

Table 4.7 Comparison of long-range coupling constants to methylenes in S^3 HMBC *homo*. Compared to the values extracted to methines from Table 4.2. [a]Theoretical coupling constants calculated as in Table 4.2. *Right*: The extraction of j_1 and j_2 from the 1D slices of cross-peaks

(C) H1	H2	j_1 [Hz]	j_2 [Hz]	J_1 [Hz]	J_2 [Hz]	$J_{methine}$ [Hz]	Other [Hz]	
4a/b	3	20.3	7.5	13.9	6.4	14.0, 6.3	$^{2}J_{CH}$ = 9.4	
11a/b	12	12.2	6.8	9.5	2.7	8.5, 3.4	$^{2}J_{CH}$ = 1.4, $^{3}J_{HH}$ = 2.5	
15a/b	16	5.7	3.3	4.5	1.2	4.0, 2.2	$^{2}J_{CH}$ = 1.3	
23a/b	22	13.2	0.4	6.8	6.4	7.0, 6.1	$^{2}J_{CH}$ = 6.8, $^{4}J_{HH}$ = 4.0	
17a/b	18a	10.1	2.6	6.3	3.7	7.5, 0.5[a]	$^{2}J_{CH}$ = 1.0, $^{2}J_{HH}$ = 8.8	
17a/b	18b	19.1	1.3	10.2	8.9	12.1, 6.3[a]	$^{2}J_{CH}$ = 3.1, $^{2}J_{HH}$ = 10.1	
18a/b	17a/b	16.2	4.8	10.5	5.7	–	–	

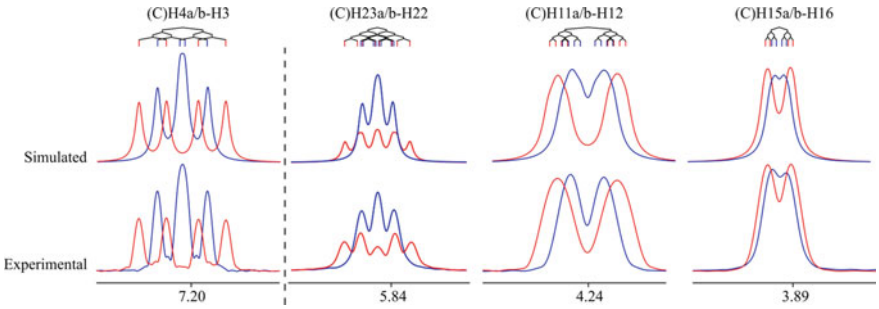


Fig. 4.20 Experimental and simulated multiplicity patterns for vinyl acetate (left, 220 mM) and strychnine (right, 300 mM). The simulated peaks are all obtained from a manual fitting in MestreNova. Coupling trees are indicative. Coupling constants found in Table 4.7

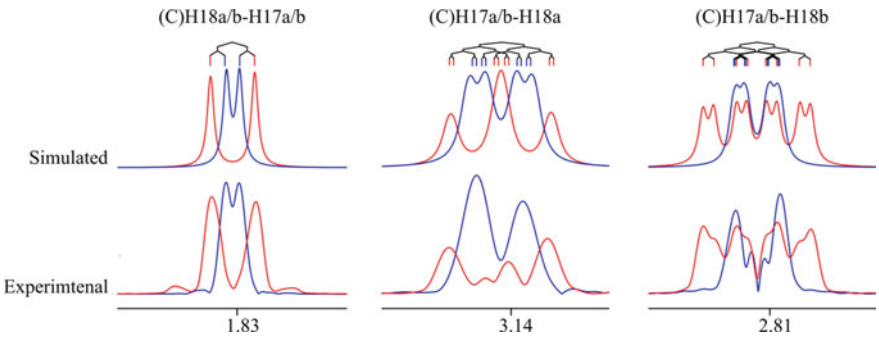


Fig. 4.21 Experimental and simulated multiplicity patterns from strychnine (300 mM). The simulated peaks are all obtained from a manual fitting of coupling constants in MestreNova. Coupling trees are indicative. Coupling constants found in Table 4.7

seen on the left for vinyl acetate. As soon as the coupling constants became smaller and/or more equal, the extractions were much more challenging (right). Especially when more passive coupling constants were added to the in-phase coupling patterns, complications ensued.

It was also possible to extract coupling constants between the 17 and 18 protons, as seen in Fig. 4.21. The theoretically most complicated peak that includes 17a/b and 18a/b was seemingly the most simple as the ^{2}J methylene coupling constants were not observed and the coupling constants seemed to average.

In conclusion, while it was possible to extract coupling constants to methylenes, the results should be considered approximates if using Eq. (4.12). More elaborate simulations of the spectra, e.g. by the program Spinach [67], are expected to yield better results, but was not performed.

4.5 Conclusion

Two new pulse sequences were introduced; the S^3 HMBC *homo* and the S^3 HMBC *hetero* for the determination of long-range homo- and heteronuclear coupling constants respectively, including size and sign. The pulse sequences were tested on the alkaloid strychnine and resulted in excellent fit of the experimental data to theoretical and reported coupling constants alike.

4.6 Experimental

The experimental setup, with relevant information is included in the various tables and figures throughout this chapter and standard pulse sequence setups are found in Appendix A1.

Acquisition

For strychnine the setup of the S^3 HMBC experiments were: ns = 16, ss = 32, si = 4096, ni = 256 J_{min} = 124 Hz, J_{max} = 169 Hz, τ = 3.4 ms, Δ = 65 ms or 62.5 ms.

For vinyl acetate the setup of the S^3 HMBC experiments were: ns = 16, ss = 32, si = 4096, ni = 256 J_{min} = 130 Hz, J_{max} = 190 Hz, τ = 3.1 ms, Δ = 62.5 ms.

5 mm tubes were used throughout (0.5 mL). For all compounds the $^1J_{CH}$-coupling constants were determined prior to acquisition to input minimum and maximum $^1J_{CH}$-coupling constants into the setup of the S^3 HMBC experiments.

Simulations

Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations [68], using the program MacroModel version 10.8 [69, 70]. The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization [69, 71].

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations [72]. Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

References

1. L. Kjaerulff, A.J. Benie, C. Hoeck, C.H. Gotfredsen, O.W. Sørensen, J. Magn. Reson. **263**, 101–107 (2016)
2. C. Hoeck, C.H. Gotfredsen, O.W. Sørensen, J. Magn. Reson. **275**, 68–72 (2017)
3. L. Kjærulff, *NMR Structural Studies of Oligosaccharides and Other Natural Products* (Technical University of Denmark, 2014)

4. C.A.G. Haasnoot, F.A.A.M. DeLeeuw, C. Altona, Tetrahedron **36**, 2783–2792 (1979)
5. S. Karplus, M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **69**, 3204–3206 (1972)
6. Z. Gattin, J. Zaugg, W.F. van Gunsteren, ChemPhysChem **11**, 830–835 (2010)
7. A. Wu, D. Cremer, Int. J. Mol. Sci. **4**, 158–192 (2003)
8. G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, Chem. Rev. **107**, 3744–3779 (2007)
9. T. Rundlöf, A. Kjellberg, C. Damberg, T. Nishida, G. Widmalm, Magn. Reson. Chem. **36**, 839–847 (1998)
10. P.E. Hansen, Prog. Nucl. Magn. Reson. Spectrosc. **14**, 175–296 (1981)
11. I. Tvaroska, M. Hricovíni, E. Petráková, Carbohydr. Res. **189**, 359–362 (1989)
12. M. Barfield, J. Chem. Phys. **41**, 3825–3832 (1964)
13. M. Barfield, J. Am. Chem. Soc. **93**, 1066–1071 (1971)
14. T. Bally, P.R. Rablen, J. Org. Chem. **76**, 4818–4830 (2011)
15. F. Cen-Pacheco, J. Rodríguez, M. Norte, J.J. Fernández, A. Hernández Daranas, Chem. Eur. J. **19**, 8525–8532 (2013)
16. P. García, M. Martín-Pastor, Á.R. De Lera, R. Álvarez, Magn. Reson. Chem. **48**, 543–549 (2010)
17. H. Günther, *NMR Spectroscopy*, 3rd edn. (Wiley-VCH, Weinheim, 2013)
18. J.M. Schmidt, O.W. Sørensen, R.R. Ernst, J. Magn. Reson. Ser. A **109**, 80–89 (1994)
19. L. Mahi, J.C. Duplan, Magn. Reson. Chem. **35**, 379–383 (1997)
20. M. Barfield, B. Chakrabarti, Chem. Rev. **69**, 757–778 (1969)
21. T. Parella, J.F. Espinosa, Prog. Nucl. Magn. Reson. Spectrosc. **73**, 17–55 (2013)
22. B.L. Marquez, W.H. Gerwick, R.T. Williamson, Magn. Reson. Chem. **39**, 499–530 (2001)
23. A.R. Katritzky, N.G. Akhmedov, A. Güven, J. Doskocz, R.G. Akhmedova, S. Majumder, C. D. Hall, J. Mol. Struct. **787**, 131–147 (2006)
24. M.H.A. Elgamal, N.H. Elewa, E.A.M. Elkhrisy, H. Duddeck, Phytochemistry **18**, 139–143 (1978)
25. M. Hansen, H.J. Jakobsen, J. Magn. Reson. **20**, 520–529 (1975)
26. G. Zhu, A. Renwick, A. Bax, J. Magn. Reson. Ser. A **110**, 257–261 (1994)
27. G. Zhu, A. Bax, J. Magn. Reson. Ser. A **104**, 353–357 (1993)
28. V.V. Krishnamurthy, J. Magn. Reson. Ser. A **121**, 33–41 (1996)
29. C.P. Butts, B. Heise, G. Tatolo, Org. Lett. **14**, 3256–3259 (2012)
30. A. Meissner, O. Sørensen, Magn. Reson. Chem. **39**, 49–52 (2001)
31. W. Kozminski, D. Nanz, J. Magn. Reson. **142**, 294–299 (2000)
32. J. Saurí, P. Nolis, L. Castañar, A. Virgili, T. Parella, J. Magn. Reson. **224**, 101–106 (2012)
33. K.E. Kövér, O. Prakash, V.J. Hruby, J. Magn. Reson. Ser. A **103**, 92–96 (1993)
34. K.E. Kövér, V.J. Hruby, D. Uhrín, J. Magn. Reson. **129**, 125–129 (1997)
35. W. Kozminski, J. Magn. Reson. **137**, 408–412 (1999)
36. S. Gil, J.F. Espinosa, T. Parella, J. Magn. Reson. **207**, 312–321 (2010)
37. S. Gil, J.F. Espinosa, T. Parella, J. Magn. Reson. **213**, 145–150 (2011)
38. K. Ding, Magn. Reson. Chem. **38**, 321–323 (2000)
39. J. Saurí, T. Parella, Magn. Reson. Chem. **50**, 717–721 (2012)
40. M. Kurz, P. Schmieder, H. Kessler, Angew. Chem. Int. Ed. Engl. **30**, 1329–1331 (1991)
41. M.D. Sørensen, A. Meissner, O.W. Sørensen, J. Magn. Reson. **137**, 237–242 (1999)
42. N.T. Nyberg, J.Ø. Duus, O.W. Sørensen, J. Am. Chem. Soc. **127**, 6154–6155 (2005)
43. N.T. Nyberg, J.Ø. Duus, O.W. Sørensen, Magn. Reson. Chem. **43**, 971–974 (2005)
44. A.J. Benie, O.W. Sørensen, Magn. Reson. Chem. **44**, 739–743 (2006)
45. A.J. Benie, O.W. Sørensen, J. Magn. Reson. **184**, 315–321 (2007)
46. A. Meissner, J.Ø. Duus, O.W. Sørensen, J. Biomol. NMR **10**, 89–94 (1997)
47. M.D. Sørensen, A. Meissner, O.W. Sørensen, J. Biomol. NMR **10**, 181–186 (1997)
48. A. Meissner, J.Ø. Duus, O.W. Sørensen, J. Magn. Reson. **128**, 92–97 (1997)
49. A. Meissner, T. Schulte-herbrüggen, J. Briand, A. Meissner, T. Schulte-herbruggen, O.W. Srensen, Mol. Phys. **95**, 1137–1142 (1998)
50. J. Keeler, *Understanding NMR Spectroscopy*, 2nd edn (Wiley, 2010)
51. O.W. Sørensen, Prog. Nucl. Magn. Reson. Spectrosc. **21**, 503–569 (1989)
52. K.M. Crecely, R.W. Crecely, J.H. Goldstein, J. Mol. Spectrosc. **37**, 252–259 (1971)
53. A. Meissner, T. Schulte-Herbrüggen, O.W. Sørensen, J. Am. Chem. Soc. **120**, 7989–7990 (1998)

54. A. Bagno, F. Rastrelli, G. Saielli, Chem. Eur. J. **12**, 5514–5525 (2006)
55. M.W. Lodewyk, M.R. Siebert, D.J. Tantillo, Chem. Rev. **112**, 1839–1862 (2012)
56. K.W. Wiitala, C.J. Cramer, T.R. Hoye, Magn. Reson. Chem. **45**, 819–829 (2007)
57. R. Jain, T. Bally, P.R. Rablen, J. Org. Chem. **74**, 4017–4023 (2009)
58. C. Adamo, V. Barone, J. Chem. Phys. **108**, 664–675 (1998)
59. M. Schmidt, F. Reinscheid, H. Sun, H. Abromeit, G.K.E. Scriba, F.D. Sönnichsen, M. John, U.M. Reinscheid, Eur. J. Org. Chem. 1147–1150 (2014)
60. R.A.E. Edden, J. Keeler, J. Magn. Reson. **166**, 53–68 (2004)
61. V. Blechta, F. del Río-Portilla, R. Freeman, Magn. Reson. Chem. **32**, 134–137 (1994)
62. E. Hatcher, E. Säwén, G. Widmalm, A.D. Mackerell Jr., J. Phys. Chem. B **115**, 597–608 (2011)
63. A. Kjellberg, G. Widmalm, Biopolymers **50**, 391–399 (1999)
64. M. Martin-Pastor, C.A. Bush, J. Biomol. NMR **19**, 125–139 (2001)
65. R. Pendrill, E. Säwén, G. Widmalm, J. Phys. Chem. B **117**, 14709–14722 (2013)
66. S.G. Smith, J.M. Goodman, J. Am. Chem. Soc. **132**, 12946–12959 (2010)
67. H.J. Hogben, M. Krzystyniak, G.T.P. Charnock, P.J. Hore, I. Kuprov, J. Magn. Reson. **208**, 179–194 (2011)
68. Maestro (Schrödinger, LLC, New York, 2015)
69. Schrödinger (LLC, New York, NY 2016)
70. Desmond (D. E. Shaw Research, New York 2015)
71. G. Chang, W.C. Guida, W.C. Still, J. Am. Chem. Soc. **111**, 4379–4386 (1989)
72. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K. N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V. G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in *Gaussian 09, Revision B.01* (Gaussian Inc., Wallingford, CT, 2010)

Chapter 5
Theory—3D Structural Information from NMR, Part 2

The following chapters all concern the use of residual dipolar coupling constants to obtain 3D structural information. Theory on the subject is thus presented below.

5.1 Residual Dipolar Coupling

Residual dipolar coupling constants (RDCs) have gradually made the move from macromolecules to small molecules over the last couple of decades. In 2000, Courtieu et al. reported enantiodiscriminating properties of poly-γ-benzyl-L-glutamate (PBLG) for the very small molecule 1-chloropropan-2-ol (CP), one of the methods being RDC analysis [1]. In the following years, Thiele and coworkers showed the possibility of solving the pro-chirality of the methylene groups in strychnine, illustrating the possibility of discerning stereoisomers of small molecules [2, 3]. Later again, the groups of e.g. Thiele and Reggelin started to investigate the enantiodiscriminating properties of PBLG and other, novel, liquid crystals [4–7]. In parallel, pioneered by the work of Kessler et al. in 2004, an investigation into the usage of swollen polymers as alignment media for small molecules began, utilizing an easier scalability of experimental RDCs as well as a weaker degree of alignment [8, 9].

The basis of the residual dipolar couplings is the direct dipolar coupling constant (D). Akin to the NOE relaxation, D-couplings operate through space via magnetic dipole-dipole interactions from the magnetic fields generated by the spins [10]. In contrast to NOEs and J-coupling constants, which are products of interactions between the nuclear spins, dipolar couplings depend on the direct influence of the external magnetic field on internuclear vectors [10–12]. D-coupling constants depend on the gyromagnetic ratio (γ) of the involved nuclei, the cubed length of the inter-nuclear vector (r) and the angle of said vector to the magnetic field (θ).[1] The

[1]Note that θ changes definition from dihedral angle to angle with magnetic field dependent on whether J-coupling constants or RDCs are the current subject.

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_5

other constants are the vacuum permeability constant (μ_0) and the reduced Planck constant ($\hbar$). The relevant formula is given in Eqs. (5.1) and (5.2) [10–13].

$$\widehat{\mathcal{H}}_{IS}^{DD,full} = -\frac{\mu_0}{8\pi^2}\frac{\hbar\gamma_I\gamma_S}{r_{IS}^3}\left(3\left[\widehat{I}_I \cdot e_{IS}\right]\left[\widehat{I}_S \cdot e_{IS}\right] - \widehat{I}_I \cdot \widehat{I}_S\right) \tag{5.1}$$

where e_{IS} is a vector through the two involved nuclei defined so that $e_{IS} \cdot e_{IS} = 1$. For two spin ½ nuclei and using that the scalar product of the internuclear vector and the magnetic field vector is equal to cos θ_{IS}, Eq. (5.1) may be rewritten to Eq. (5.2).[2]

$$D_{IS} = -\frac{\hbar\gamma_I\gamma_S\mu_0}{16\pi^2}\left\langle\frac{1}{r_{IS}^3}\left(3\cos^2\theta_{IS} - 1\right)\right\rangle = -D_{max}\left\langle\frac{1}{r_{IS}^3}\left(3\cos^2\theta_{IS} - 1\right)\right\rangle \tag{5.2}$$

In anisotropic liquid conditions the observed angle is an average of the dipolar coupling over all orientations. The dependence on the angle to the magnetic field may be used in determining the 3D structures of molecules as it will establish angular knowledge of internuclear vectors even far apart in a structure, as exemplified in Fig. 5.1, which may be utilized by back-calculation and fitting of experimental data to a proposed structure.

This can make it possible to correlate clusters of stereo-centers in a given structure that are not connected by protons in close proximity, which is impossible using NOEs and *J*-coupling constants due to their local nature.

In isotropic tumbling, i.e. where the molecules are freely moving in solution, the term $(3\cos^2\theta_{IS} - 1)$ averages to zero and no RDCs are observed, taking into account that all orientations have equal probability, in Eq. (5.3) [10].

$$D_{IS} = \int_0^{\pi} \sin\theta_{IS}\left(3\cos^2\theta_{IS} - 1\right)d\theta_{IS} = 0 \tag{5.3}$$

This implies that *D*-coupling constants may not be observed in isotropic liquid NMR spectroscopy as signal splittings, and a surplus of a given alignment to the magnetic field needs to be introduced making the sample anisotropic. The anisotropic sample is here characterized by a preferred molecular orientation as well as rapid translational motions of the molecules [10]. The result is that intramolecular dipoles give rise to dipolar coupling constants, while intermolecular dipole interactions will still average to zero as a result of translational motions, at least to a good approximation and for short (and thereby larger observed) intermolecular dipole interactions [10].

Were all molecules equally aligned, the full dipolar coupling would be observed and the constants from Eq. (5.2) yield *D*-coupling constants with a magnitude of tens of thousands Hz for $^1D_{CH}$ ($D_{max}/r^{-3} \sim 11{,}350$ Hz). This would make them

[2]Compared to the notation by Levitt, the dipolar coupling is divided by 2π to gain values in Hz instead to radians/s.

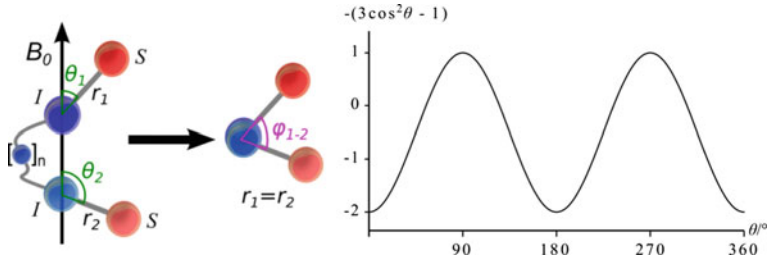


Fig. 5.1 Left: Correlation of the angle between two internuclear vectors (φ) from the angle of the internuclear vectors to the magnetic field (θ). Right: The relative dependence of RDCs on the angle θ

hard to measure, and the sign next to impossible to determine, when comparing to $^1J_{CH}$ (typically +100 to +200 Hz) [12, 13]. Of course other problems such as intermolecular dipolar couplings and broader line-shapes due to a multitude of large long-range dipolar couplings etc. may also be envisioned. Thus the alignment needs to create a smaller surplus of one orientation, with the observed value being the residual dipolar coupling (RDC), preferably in the magnitude of tens of Hz which translates to a $\sim 10^{-3}$ fraction of alignment. This is in practice achieved by using alignment media in liquid solutions which align only an appropriate amount of molecules leading to average coupling constants of the desired maximum absolute values. Note that D will refer to a residual dipolar coupling constant in the rest of this thesis.

In an anisotropic sample the D-coupling constants are extracted from the difference between the observed total coupling (T) and the J-coupling from isotropic tumbling as in Eq. (5.4).

$$D = T - J \tag{5.4}$$

The sum of dipolar interactions is observed due to identical form of the respective Hamiltonians of the spin parts of J and D. The J-coupling constants are considered unaltered as they are almost independent of the orientation to the magnetic field [10]. The anisotropic part of the J-coupling (termed the J-anisotropy) is small and usually ignored, even in solids NMR [10]. The line-splitting is thus a summation of the two dipolar coupling contributions, and it should be noted that also $T = J + 2D$ is used in the literature [11]. While this will of course alter the observed RDCs, it will not influence the conclusions from back-calculation of RDCs from 3D structures, since all RDCs are scaled equally [11]. An example of an extraction is illustrated in Fig. 5.2. The J- and T-coupling constants are determined by displacement of peaks to maximum overlay throughout the thesis.

To investigate the degree of alignment, the quadrupolar splitting of deuterium nuclei in the deuterated NMR solvents is often used, see Fig. 5.3. From the quadrupolar splitting it may be e.g. evaluated whether the alignment is stable across long experiments, and there is a correlation between the quadrupolar splitting and

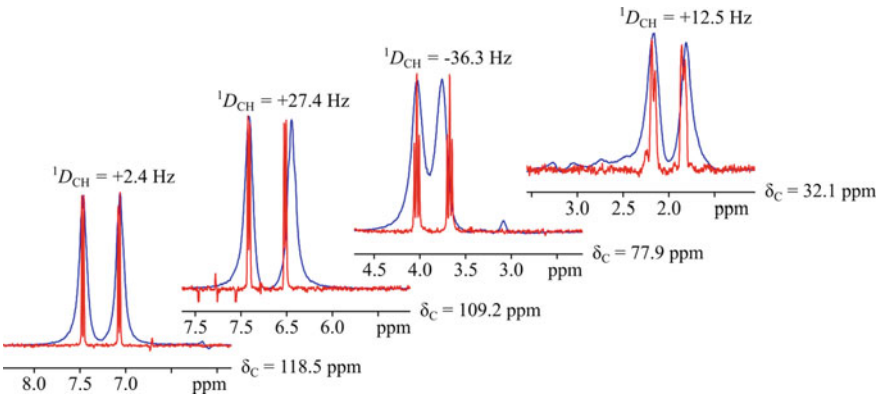


Fig. 5.2 1D slices of rows (δ_C indicated) extracted from a CLIP-HSQC of reserpine (see Sect. 8.6.2) under isotropic conditions yielding *J*-coupling constants (red) and anisotropic conditions yielding total coupling constant (blue). Aligned in PDMAA (Polymer 8.2). The values of the RDCs are shown above the doublets

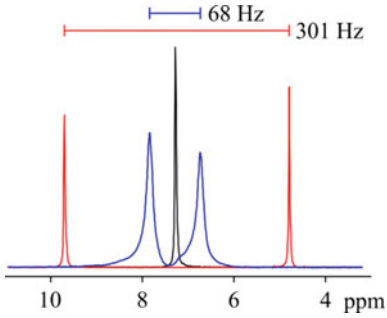


Fig. 5.3 Examples of the ^{2}H quadrupolar splitting for $CDCl_3$ of an isotropic sample (black), and aligned samples in the liquid crystal PBLG (red) and the stretched polymer PDMAA (blue). The size of the splitting is indicated above at 400 MHz (^{1}H frequency)

the size of the observed RDCs for many alignment media [14, 15]. The linear correlation between the size of RDCs and ^{2}H-splitting is compound and alignment media dependent. Not all alignment media give rise to observable quadrupolar splittings, especially the more polar solvents, such as DMSO, seem to exhibit smaller splittings, which may be hidden in the linewidth [16].

The measured RDCs are compared to 3D structures of the investigated molecule by back-calculation, where experimental data is compared to already rendered 3D structure(s). For bio-macromolecules, RDCs are also used in the optimization and simulation of structures, but the method has not caught on within the small molecule community and will not be elaborated on [17]. The preferred back-calculation of RDCs from 3D structures for small molecules is singular value decomposition (SVD) [18–20], which has proven reliable and generally applicable for the structures frequently used in the literature. They generally all share a common feature: They are rigid. The requirement of rigidity in the structure is a major problem for general usage of RDCs, which is addressed in Chap. 8, along

with a more detailed introduction to SVD. Other methods have also been proposed such as prediction of alignment based on e.g. shape of the molecule, but these are generally less utilized in the small molecule literature [19–21].

The used experiment for determining $^1J_{CH}$ and $^1T_{CH}$ was generally the CLIP experiment, which exhibit the line splitting of the coupling constant in the direct (F2) dimension [22, 23]. Other groups have shown that using the indirect (F1) dimension may be advantageous due to e.g. that homonuclear coupling constants will not influence the line splitting in the indirect dimension, resulting in potentially sharper lines [11, 24]. The experimental time is increased though due to the lower resolution in F1, even when *J*-scaling is used. The author experimented slight differences in the extracted coupling constants between the F1 and F2 experiments, and it was thought advantageous to keep to one type of experiment throughout.

5.1.1 Alignment Media

The available alignment media for small molecules in organic solvents are generally divided into two classes: Liquid crystals (LCs) and stretched polymers (SPs) [11, 12]. The latter class is also termed strain-induced alignment in a gel (SAG) effect alignment media. The amount of alignment media is different compared to those for macromolecules due to a need for applicability in organic solvent. Thus e.g. filamentous phages, rod-shaped cellulose particles or charged acrylamide gels are not generally applicable [25]. The interaction with the alignment media is generally thought to be solely steric, unless a charged alignment media is used [19, 26]. A list of alignment media and the solvents in which they may be used is found in Table 5.1 [13].

Multiple alignment media are available, and an outline of advantages and limitations of LCs and SPs is given below, as a detailed discussion of all would be beyond the scope of this text. General consideration focused on enantiodiscrimination will also be touched upon as an introduction to Chap. 7. The work concerning enantiodiscrimination of organic molecules has primarily been centered on LCs [4, 6, 7, 14, 28]. These include PBLG, polyacetylenes and polyguanidines. For comparison only a single stretched radical-initiated crosslinked polymer has been published [26], while crosslinked amino acid based stretched polymers such as gelatin have also been shown to possess enantiodiscriminative properties [35].

5.1.2 Liquid Crystals

LCs for the application as small molecular alignment media are typically made of long polymeric single strands that possess a secondary structure. The LCs are normally helical but sheet structures have been reported in the form of graphene

Table 5.1 Examples of liquid crystal (LC) and stretched polymers (SP) alignment media and the most utilized solvents used for alignment in the media

	Solvent	Chiral	References
LC			
PBLG	$CDCl_3$	Yes	[1, 14, 27]
PELG	$CDCl_3$	Yes	[3]
Guanidines	$CDCl_3$	Yes	[28]
Acetylenes	$CDCl_3$	Yes	[4, 7]
Graphene	DMSO	No	[29, 30]
SP			
PS	$CDCl_3$	No	[8, 31]
PDMS	$CDCl_3$	No	[32]
PMMA	$CDCl_3$	No	[9]
PVAc	$CDCl_3$ CD_3OD DMSO	No	[16]
PH/PAA	D_2O DMSO	No	[33, 34]
PDMAA/(R/S)-APhES	DMSO	Yes	[26]
Gelatin	D_2O	Yes	[35]

Inspired by Thiele [13]

layers [14, 29]. The most prominent member of the class is arguably PBLG, and this is also the only LC the author has worked with.

The anisotropy is introduced by dissolving of the LC in the given solvent [5, 14, 36]. By repeatedly transferring the solution from one end of the NMR tube to the other, e.g. by use of rotor, the solution is homogenized. The homogeneity along the length of the NMR tube is verified by measuring the quadrupolar splitting of the deuterium signal. The anisotropy is induced by interactions of the molecules and the vertically aligned helices as illustrated in Fig. 5.4.

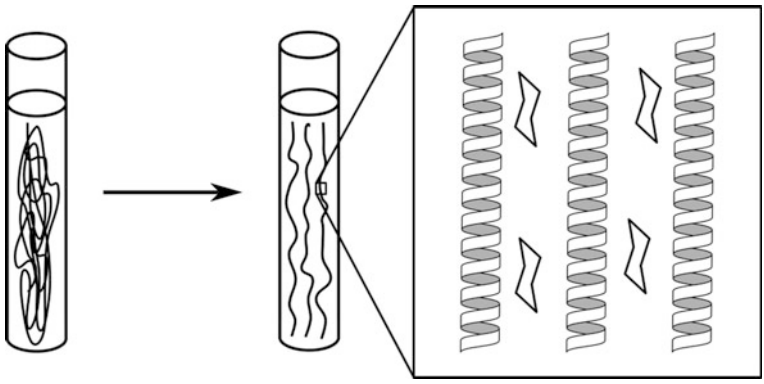


Fig. 5.4 Illustration of the alignment of molecules in liquid crystals (LC), based on PBLG. The induced anisotropy is indicated

The main disadvantage of the LCs is that they are hardly scalable, though different molecular weight PBLG was shown to result in differences in the alignment strength [14]. Also the alignment is strong, leading to larger RDCs, which may lead to problems and limit the scope of usage [14]. An advantage is that many LCs are readily available from vendors, the sample preparation of the alignment media is easy and the line shapes of aligned compounds are, based on experimental experience with PBLG of the author, sharper than when using stretched polymers. Also, most of the LCs are inherently chiral and enantiomeric pure, leading to better discrimination between enantiomers [4, 7, 14, 28].

While PBLG is probably the most utilized LC for RDC extraction, the enantiodiscriminative prowess of the polymer is not great compared to other reported LCs. It was shown though, that the alignment of (+)- and (−)-IPC, and thus the observed RDCs, was affected by the amount of PBLG versus PBDG (the enantiomeric polymer), and that differentiation was indeed possible [5, 14]. Also cross-linked PBLG (which could in theory be considered a stretched polymer) was shown to differentiate IPC [6]. It was actually reported that the differentiation was increased by crosslinking the LC.

Other liquid crystals have been published with greater enantiodiscriminative properties, see Fig. 5.5 for examples. The first was a polyguanidine which showed much differences in the orientation of the enantiomers of IPC [28]. This was followed by two polyacetylenes which also showed great differentiation [4, 7]. Interestingly, one of the polyacetylenes was reported to exhibit enantiodiscriminative properties even after the temperature was varied to a point where circular dichroism showed no left- or right-handedness of the helical backbone, suggesting that the enantiodiscriminative properties may in part be due to side-chain chirality [7]. This was surprising since the enantiodiscriminative properties of the LCs were thought to originate mostly from their helical structure, while this finding suggest an interaction to the chiral center presented at the side-chain [4, 5, 7, 28].

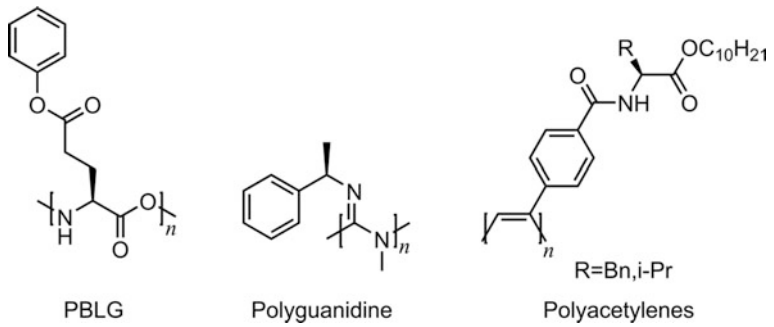


Fig. 5.5 Examples of chiral LCs from the literature. All give rise to helixes in solution [4, 5, 7, 14, 28]

5.1.3 *Stretched Polymers*

Stretched polymers cover crosslinked polymers that swell in the used solvents. The term stretched polymers is fitting especially for polymers which are mechanically scaled using an amply named stretching apparatus [15, 37–39]. The stretching apparatus consists of a rubber tube and clams which uniformly stretches the gel inside the tube, and may be used to tune the alignment. Rubber tubes for different solvents have been introduced to allow the usage of e.g. chloroform [15].

Another method of swelling the polymer is by synthesizing polymers of a size which limits swelling to be either radially or vertically in the NMR tube. The polymer stick is then swollen in a suitable solvent and the result is an anisotropic environment as illustrated in Fig. 5.6 [9, 16]. The anisotropic environment may be simplified to the steric interactions with the irregular distribution of interaction surfaces with either horizontal or vertical surface in excess, formed once the polymer is swollen [9]. In practice, the polymers were synthesized in either 5 mm (vertical swelling) or 3 mm (radial swelling) NMR tubes, and subsequently swollen in 5 mm NMR tubes [9].

The alignment strength may often not be altered once swollen, but may be tuned by the amount of cross-linker used during synthesis and the size of chosen polymer stick [9, 16]. It should be noted that the synthesis of these polymers seems much easier than that of e.g. PBLG (though the latter was not attempted by the author) [14], and it is hardly problematic that scaling depends on cross-linker content. The vertical swelling was first introduced, but the time needed to reach full swelling is generally long, up to several weeks [9]. Thus the concept of radial swelling of thin polymers which were vertically restricted was introduced [9]. Some of the polymers may be shifted from constricted to un-constricted multiple times while other do not possess this property and should not be relaxed after swelling [9, 26].

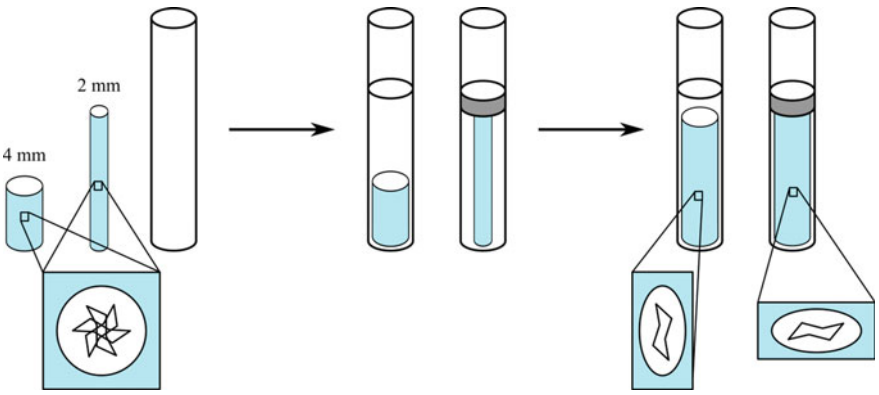
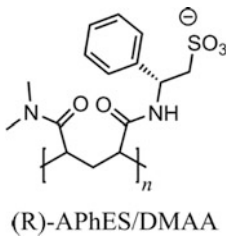


Fig. 5.6 Schematic representation of the alignment of molecules in stretched polymers, either swollen vertically or radially in an appropriate solvent. The anisotropy of the polymers upon swelling is indicated



(R)-APhES/DMAA

Fig. 5.7 Example of a chiral stretched polymer from the literature, shown to work only by interactions to a charged amine moiety of the analyte [26]

The amount of stretched polymers which will differentiate enantiomers is, as mentioned previously, lower than the number for LCs. Polymers are based on crosslinked amino acids, such as gelatin and the underlying biopolymer of gelatin, collagen, were shown to differentiate diastereomers in aqueous solution [35, 39, 40]. A more classic (i.e. radical initiated) polymer based on a chiral monomer alike to the achiral AMPS monomer, (*R*/*S*)-2-acrylamido-2-methylpropane-1-sulfonic acid (APhES), and the achiral monomer (*N,N*)-dimethylacrylamide (1:1) was later introduced, which was shown to differentiate compounds that possess a charged amine group under acidic conditions in DMSO [26, 33] (Fig. 5.7).

5.1.4 Evaluation of RDCs

To evaluate the experimental RDCs in comparison to the given 3D structure(s), the *Q*-factor is utilized as given in Eq. (5.5) [19, 20].

$$Q = \sqrt{\frac{\sum \left(D_i^{exp} - D_i^{calc}\right)^2}{\sum D_i^{exp\,2}}} \tag{5.5}$$

The *Q*-factor approaches zero when the difference between experimental and back-calculated values approaches zero and the size of the extracted RDCs is included in the calculations, which leads to easy comparison of fits between dataset of different degrees of alignment.

5.1.5 Parameters Used to Investigate Enantiodiscrimination

Some different parameters have been proposed in order to compare the alignment of different enantiomers. One is the 5D or generalized cosine β (GCB) angle, which directly compares the alignment tensors from e.g. SVD fittings, given in Eq. (5.6) as the normalized scalar product between the alignment tensors (see Sect. 8.1 for the theory of SVD and alignment tensors) [7, 41].

$$GCB = \frac{\langle A^1 | A^2 \rangle}{\langle A^1 | A^1 \rangle^{1/2} \langle A^2 | A^2 \rangle^{1/2}}$$
$$\langle A^1 | A^2 \rangle = \sum_{i,j} A^1_{ij} A^2_{ij} \tag{5.6}$$

An F-test has also been utilized to check whether the actual RDC values are statistically different, using the null hypothesis that there is no significant difference (standard for the F-test) [26]. The F-test is used to compare variance of the data and the difference in variance between datasets of different and identical enantiomers are compared. It should be noted that the F-test is generally quite sensitive toward non-normality and it is not used in this thesis.

References

1. M. Sarfati, P. Lesot, D. Merlet, J. Courtieu, Chem. Commun. 2069–2081 (2000)
2. C.M. Thiele, S. Berger, Org. Lett. **5**, 705–708 (2003)
3. C.M. Thiele, J. Org. Chem. **69**, 7403–7413 (2004)
4. N.C. Meyer, A. Krupp, V. Schmidts, C.M. Thiele, M. Reggelin, Angew. Chemie Int. Ed. **51**, 8334–8338 (2012)
5. A. Marx, V. Schmidts, C.M. Thiele, Magn. Reson. Chem. **47**, 734–740 (2009)
6. T. Montag, C.M. Thiele, Chem. Eur. J. **19**, 2271–2274 (2013)
7. A. Krupp, M. Reggelin, Magn. Reson. Chem. **50**, S45–S52 (2012)
8. B. Luy, K. Kobzar, H. Kessler, Angew. Chemie Int. Ed. **43**, 1092–1094 (2004)
9. C. Gayathri, N.V. Tsarevsky, R.R. Gil, Chem. Eur. J. **16**, 3622–3626 (2010)
10. M.H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd edn. (John Wiley & Sons Ltd, Chichester, 2008)
11. C.M. Thiele, Eur. J. Org. Chem. 5673–5685 (2008)
12. G. Kummerlöwe, B. Luy, Trends Anal. Chem. **28**, 483–493 (2009)
13. C.M. Thiele, Concepts Magn. Reson. Part A **30A**, 65–80 (2007)
14. A. Marx, C. Thiele, Chem. Eur. J. **15**, 254–260 (2009)
15. G. Kummerlöwe, E.F. McCord, S.F. Cheatham, S. Niss, R.W. Schnell, B. Luy, Chem. Eur. J. **16**, 7087–7089 (2010)
16. J.C. Freudenberger, S. Knör, K. Kobzar, D. Heckmann, T. Paululat, H. Kessler, B. Luy, Angew. Chemie Int. Ed. **44**, 423–426 (2005)
17. M. Blackledge, Prog. Nucl. Magn. Reson. Spectrosc. **46**, 23–61 (2005)
18. J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, J. Magn. Reson. **138**, 334–342 (1999)
19. M. Zweckstetter, Nat. Protoc. **3**, 679–690 (2008)
20. A. Navarro-Vázquez, Magn. Reson. Chem. **50**, S73–S79 (2012)
21. P. Tzvetkova, B. Luy, S. Simova, Magn. Reson. Chem. **50**, S92–S101 (2012)
22. A. Enthart, J.C. Freudenberger, J. Furrer, H. Kessler, B. Luy, J. Magn. Reson. **192**, 314–322 (2008)
23. I. Timári, L. Kaltschnee, A. Kolmer, R.W. Adams, M. Nilsson, C.M. Thiele, G.A. Morris, K. E. Kövér, J. Magn. Reson. **239**, 130–138 (2014)
24. J.D. Snider, E. Troche-Pesqueira, S.R. Woodruff, C. Gayathri, N.V. Tsarevsky, R.R. Gil, Magn. Reson. Chem. **50**, S86–S91 (2012)
25. A. Bax, A. Grishaev, Curr. Opin. Struct. Biol. **15**, 563–570 (2005)

26. M. Schmidt, H. Sun, A. Leonov, C. Griesinger, U.M. Reinscheid, Magn. Reson. Chem. **50**, S38–S44 (2012)
27. P. Lesot, M. Sarfati, J. Courtieu, Chem. Eur. J. **9**, 1724–1745 (2003)
28. L. Arnold, A. Marx, C.M. Thiele, M. Reggelin, Chem. Eur. J. **16**, 10342–10346 (2010)
29. X. Lei, Z. Xu, H. Sun, S. Wang, C. Griesinger, L. Peng, C. Gao, R.X. Tan, J. Am. Chem. Soc. **136**, 11280–11283 (2014)
30. W. Zong, G. Li, J. Cao, X. Lei, M. Hu, H. Sun, C. Griesinger, R.X. Tan, Angew. Chemie Int. Ed. **55**, 3690–3693 (2016)
31. B. Luy, K. Kobzar, S. Knör, J. Furrer, D. Heckmann, H. Kessler, J. Am. Chem. Soc. **127**, 6459–6465 (2005)
32. J.C. Freudenberger, P. Spiteller, R. Bauer, H. Kessler, B. Luy, J. Am. Chem. Soc. **126**, 14690–14691 (2004)
33. P. Haberz, J. Farjon, C. Griesinger, Angew. Chemie—Int. Ed. **44**, 427–429 (2005)
34. S. Meier, D. Häussinger, S. Grzesiek, J. Biomol. NMR **24**, 351–356 (2002)
35. K. Kobzar, H. Kessler, B. Luy, Angew. Chem. Int. Ed. Engl. **44**, 3145–3147 (2005)
36. A. Marx, B. Böttcher, C.M. Thiele, Chem. Eur. J. **16**, 1656–1663 (2010)
37. P.W. Kuchel, B.E. Chapman, N. Müller, W.A. Bubb, D.J. Philp, A.M. Torres, J. Magn. Reson. **180**, 256–265 (2006)
38. G. Kummerlöwe, F. Halbach, B. Laufer, B. Luy, Open Spectrosc. J. **2**, 29–33 (2008)
39. G. Kummerlöwe, M.U. Kiran, B. Luy, Chem. Eur. J. **15**, 12192–12195 (2009)
40. B. Luy, J. Indian Inst. Sci. **90**, 119–132 (2010)
41. J. Sass, F. Cordier, A. Hoffmann, A. Cousin, J.G. Omichinski, H. Löwen, S. Grzesiek, J. Am. Chem. Soc. **121**, 2047–2055 (1999)

Chapter 6
Determination of Long-Range Residual Dipolar Coupling Constants

A major hindrance in the scope of utilizing RDCs for structural studies of small molecules is an inherent lack of internuclear vectors to measure and compare to 3D structures. Whereas $^{1}D_{XH}$ are plentiful for macromolecules, only utilizing the one bond couplings is a limitation for small molecule RDCs. Thus it would be beneficial to increase the number of vectors by including long-range RDCs. For this purpose, the S^3 HMBC experiments may prove very useful. While SVD is reviewed in Chap. 8, it is worth mentioning that it is the most widely used back-calculation method for small organic molecules, and that a minimum of five linearly independent internuclear vectors are needed to obtain a fit between RDC data and a rigid 3D structure. If multiple moving parts of a structure need to be investigated independently, five independent internuclear vectors are needed *per moving part*, which may be a concern for small molecules [1, 2].

6.1 Utilized Alignment Media

It was not straight forward to implement the usage of the S^3 HMBC experiments for RDCs due to inherent problems with the used alignment media. This may be attributed to the change in T_2 relaxation for the small molecules in aligned samples. A decrease in T_2 time due to interactions to the solid-like alignment media leads to a rapid loss of magnetization and the longer, compared to e.g. HSQC-type experiments, S^3 HMBC experiments resulted in a lack of signal to detect. Thus the search for an alignment medium which proved advantageous was initiated. Essentially two factors were deemed the most important:

1. Alignment should be weak enough to allow sufficient tumbling so that T_2 relaxation allows signal detection
2. Alignment should be strong enough that long-range RDCs are larger than the experimental error

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_6

As one might suspect, these requirements are contradicting and some alignment media may be readily discarded.

The initial choice of alignment media illustrates this dilemma; the liquid crystal PBLG and stretched PMMA and PVAc polymers [3–5]. Spectra of IPC and strychnine from PBLG failed to obtain an *S/N* ratio where RDCs could be determined in the author's hands. This is contributed to the strong alignment of the medium, perhaps best illustrated by the large observed $^{1}D_{CH}$-coupling constants. On the contrary, PMMA and pVAc would give rise to observable RDCs, but weak alignment strength and thus low RDCs, gave rise to a large possible influence from experimental errors and low discriminative prowess. This was somewhat relieved by altering the amount of cross-linker in order to increase the size of the RDCs.

In the end, an alignment medium that best satisfied the needs was identified; a new stretched polymer made from crosslinked DMAA (PDMAA). It should be noted that a similar polymer has been published, e.g. the PH/PAA polymer where 2-(acrylamido)-2-methylpropanesulfonic acid is used in addition to DMAA (in a molar fraction of 1:1) and crosslinker [6, 7]. This is supposedly done to facilitate swelling in DMSO.

It was advantageous to swell PDMAA using the confinement method, where vertical swelling is restricted and the polymer is allowed to swell radially. PDMAA polymers swell easily in both chloroform and DMSO with observed coupling constants in the order of 20–30 Hz, and are thus very well suited for the purpose of aligning small organic molecules. The observed RDCs were easily scalable by differentiation of the amount of cross-linker in line with other stretched polymers [4]. This is advantageous, native to stretched polymers over LC, as the optimal alignment of different compounds was not found by a singular polymer constitution, and that the alignment strength differed between compounds in the same polymer, even when comparable structures were aligned.

6.2 Results—S^3 HMBC *Homo*

The S^3 HMBC *hetero* experiment was a relatively late addition to the work, and was finalized while writing the thesis. Thus, long-range heteronuclear results from aligned samples are not included and this section will focus solely on S^3 HMBC *homo*.

RDCs were extracted from spectra of aligned and isotropic samples of strychnine and IPC. The basis of the alignment tensors for the structures was $^{1}D_{CH}$-coupling constants from CLIP-HSQC spectra, since the long-range $^{n}D_{HH}$-coupling constants are inherently smaller than $^{1}D_{CH}$-coupling constant due to the r^{-3} relationship. While much focus has been on measuring long-range heteronuclear coupling constants in the case of *J*-coupling constants, [8, 9] long-range homonuclear RDCs are more interesting than their hetereonuclear counterparts due to the $\sim$4:1 ratio of the gyromagnetic ratios of ^{1}H and ^{13}C, as illustrated in Table 6.1. As a result quite long distances between protons have identical potential values as the shortest of long-range

Table 6.1 Comparison of theoretical relative *D*-coupling constants to that of $^1D_{CH}$ for different distances (*r*) calculated using Eq. (5.2)

r [Å]	D_{CH}	D_{HH}	Example
1.1	1.00	–	$r(^1D_{CH})$
1.8	0.24	0.94	$r(^2D_{HH})$
2.2	0.13	0.50	$r(^2D_{CH})$
2.6	0.06	0.24	–
3.0	0.04	0.15	–

Identical angle to the magnetic field is assumed. Examples of nuclei pairs with an internuclear vector of the used distances are given

heteronuclear RDCs ($^2D_{CH}$). While examples of homonuclear RDCs are available in the literature, they generally seem sparsely used [10–12].

The number of RDCs obtained from S^3 HMBC *homo* was generally lower than the number of *J*-coupling constants extracted from isotropic samples, possibly due to an increase in T_2 relaxation in the anisotropic samples. Another explanation could be altered long-range heteronuclear coupling constants leading to too low or high coupling constants compared to the length of the delay used in the pulse sequence. The latter might possibly be alleviated by using an array of delays, but this was not tested.

6.2.1 IPC

The structure of IPC is much used, probably due to favorable rather weak alignment in most media, the rigid structural scaffold and multiple non-parallel vectors [13, 14] (Fig. 6.1).

IPC was aligned in PMMA, using the confinement method to introduce anisotropy. The measured one-bond and long-range RDCs are given in Table 6.2. Note that most of the "long-range" coupling constants were in fact $^3D_{HH}$, which could be obtained from other experiments, though the S^3 HMBC *homo* provide an accurate and easy method of performing the measurements. Also some $^4D_{HH}$ were extracted. The correlation of the data is generally good, though the *Q*-factor increased upon including the extra RDC data.

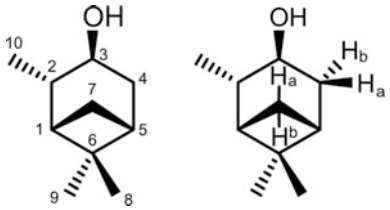


Fig. 6.1 Structure of (+)-IPC. Stereochemistry of chiral centers and diastereotopic protons indicated

Table 6.2 Experimental and back-calculated RDCs of IPC using $^{1}D_{CH}$ from CLIP-HSQC only and including $^{n}D_{HH}$ from S^3 HMBC *homo*

Nuc 1	Nuc 2	Exp (Hz)	Comp (Hz)	Exp (Hz)	Comp (Hz)
C1	H1	16.3	16.7	16.3	18.2
C2	H2	−11.6	−9.6	−11.6	−8.8
C3	H3	14.6	14.2	14.6	12.5
C4	H4a	−4.8	−4.8	−4.8	−6.1
C4	H4b	15.5	14.2	15.5	16.1
C5	H6	−7.4	−7.8	−7.4	−5.8
C7	H7a	−21.7	−23.2	−21.7	−23.8
C7	H7b	4.9	5.2	4.9	2.7
H1	H6	–	0.7	1.0	0.9
H1	H7a	–	0.2	0.5	0.0
H2	H3	–	3.7	3.7	4.0
H2	H4b	–	0.5	1.5	0.8
H3	H4a	–	1.4	1.6	1.1
H3	H4b	–	−1.7	−2.1	−1.0
H3	H6	–	−0.4	−0.3	−0.5
H4a	H6	–	−7.5	−7.0	−7.9
		Q-factor	0.076		0.147

When no experimental data (-) is given the theoretical values are calculated from the existing data. Aligned in PMMA (Polymer 6.1)

Table 6.3 *Q*-factors of back-calculations of IPC of correct assignment and with switched chirality of diastereotopic protons

Switched	*Q*-factor			
	$^{1}D_{CH}$	Ratio	$^{1}D_{CH}$ + $^{n}D_{HH}$	Ratio
Right	0.076	1.0	0.147	1.0
Dia-2	0.661	8.7	0.806	5.5
Dia-3	0.305	4.0	0.432	2.9
Dia-4	0.201	2.6	0.662	4.5
Dia-7	0.659	8.7	0.668	4.5
Dia-2 + 3	0.655	8.6	0.759	5.2
Dia-4 + 7	0.440	5.8	0.678	4.6
Dia-2 + 4+7	0.284	3.7	0.339	2.3
Dia-2 + 3+4 + 7	0.283	3.7	0.375	2.6

The additional information from the homonuclear coupling constants was used in an attempt to determine the diastereomer of the known structure. The result is found in Table 6.3 and Fig. 6.2, and the additional information led to little or no differentiation compared to only using the one bond coupling constants, as the lower values of the long-range RDCs correlate well to all 3D structures.

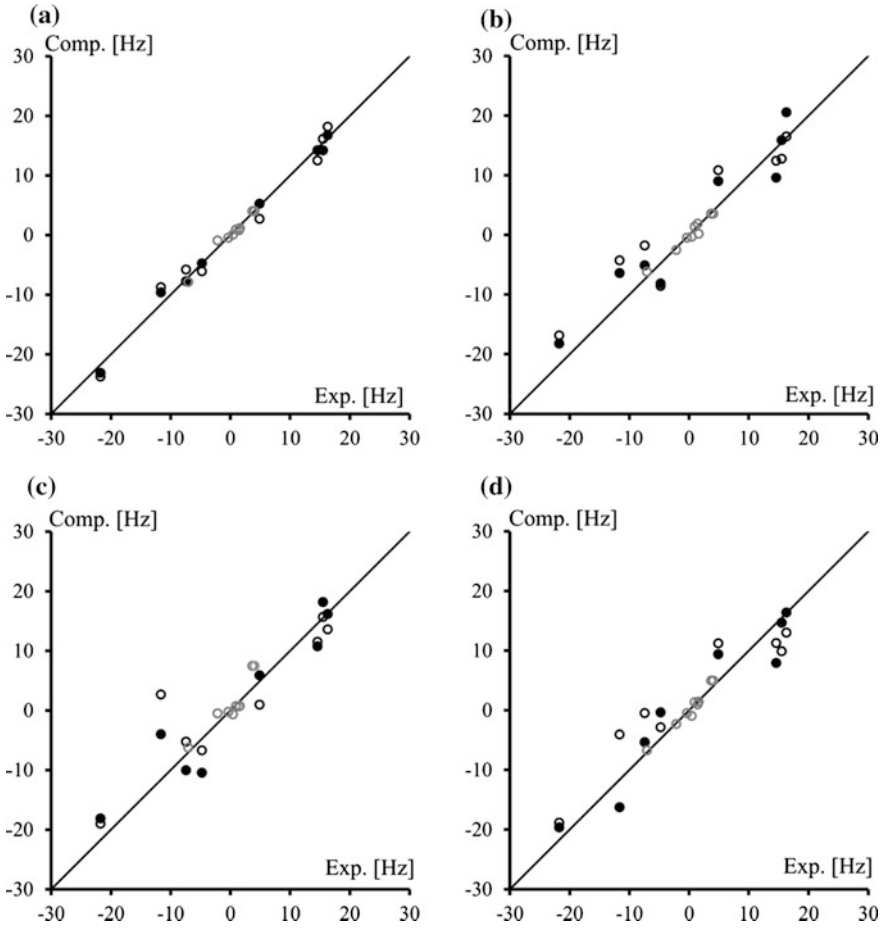


Fig. 6.2 Comparison of experimental and back-calculated RDCs of IPC. Black dots: only $^1D_{CH}$ data. Open circles: $^1D_{CH}$ and $^nD_{HH}$ data used, with $^1D_{CH}$ in black and $^nD_{HH}$ in grey. Aligned in PMMA (Polymer 6.1). **a** Correct stereochemistry, **b** 2-stereoisomer, **c** 3-stereoisomer, **d** 2- and 3-stereoisomer

6.2.2 *Strychnine*

Strychnine was also utilized to showcase the scope of the new S^3 experiments. The alignment medium was PDMAA which was identified to result in the highest number of coupling constants from S^3 HMBC *homo* versus alignment strength ratio. While PMMA gave rise to similar magnitude $^1D_{CH}$-coupling constants as DMAA, only a few $^nD_{HH}$-coupling constants could be extracted from the S^3 HMBC *homo* spectra in PMMA. Examples of extracted RDCs are given in Fig. 6.3.

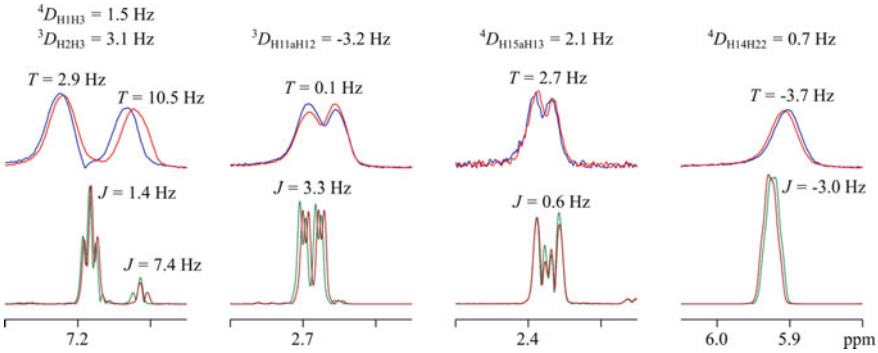


Fig. 6.3 Comparison of S^3 HMBC *homo* spectra of strychnine in $CDCl_3$ under isotropic (green and brown) and anisotropic (blue and red) conditions. Aligned by PDMAA (Polymer 6.2). The extracted coupling constants and associated RDCs are given

The general findings were similar to those of IPC, with the addition of the extra coupling constants leading to a slightly worse *Q*-factor. Many long-range coupling constants were extracted, more than doubling the available number of internuclear vectors included in the back-calculations of the data (26 $^nD_{HH}$ in addition to 18 $^1D_{CH}$) which are illustrated in Fig. 6.4, with values given in Appendix A4. While the extra internuclear vectors were not needed in the fitting of strychnine, it showed the potential of using the S^3 HMBC experiments in extracting RDCs.

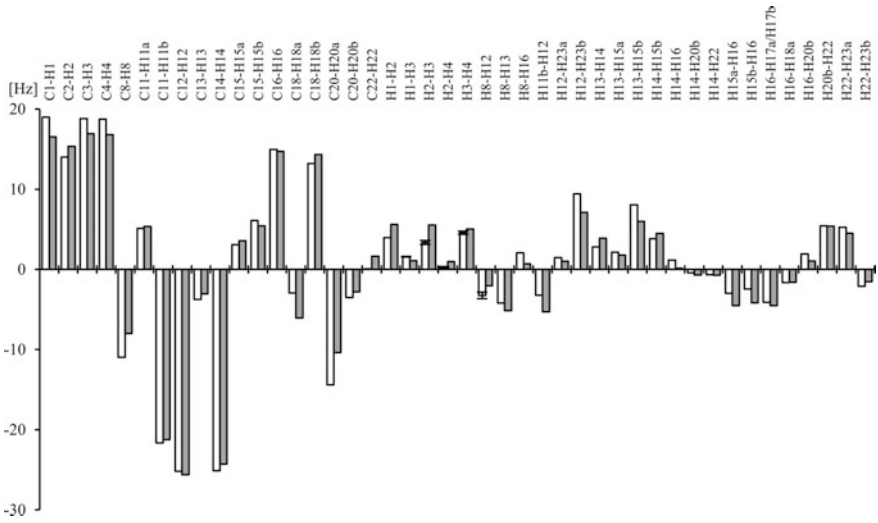


Fig. 6.4 A graphical representation of the experimental (white) and back-calculated (grey) RDCs of strychnine. When RDCs were extracted from two directions the average is given and the difference indicated with error bars. Anisotropy induced by PDMAA (Polymer 6.2)

When comparing experimental data to back-calculated values from wrongfully assigned diastereotopic protons as seen in Fig. 6.5, the best correlation between experimental and back-calculated RDCs was the correct assignment, and only the diastereomers of 15a/15b were virtually indistinguishable due to very similar RDCs. The stereochemistry of the diastereotopic protons 23a/23b could not be solved without the additional data from $^{n}D_{HH}$-coupling constants due to overlap of the resonances in the CLIP-HSQC. This is alleviated by the usage of add and subtract spectra in the S^3 HMBC *homo* as the coupling constants are not in-phase and thus the resonances are more easily differentiated. A similar distinction might be possible using the CLIP/CLAP-HSQC approach [15]. The differentiation of the diastereotopic protons was facilitated by a favorable (and lucky) orientation of the aligned molecules, where the internuclear vectors that include H23a/b had rather large and different RDCs compared to the remaining long-range RDCs.

6.3 Conclusion and Perspectives

The S^3 HMBC *homo* may be used to extract long-range homonuclear RDCs accurately and, if aligned by a suitable alignment media, in a large number. It should of course be noted here, that the chosen compounds both have a high proton to carbon ratio. The increase in available internuclear vectors may greatly impact for small molecular NMR spectroscopy, where the number of one-bond RDCs is often limited. The length of the pulse sequence may be a limitation, presumably due to a faster T_2 relaxation induced by the alignment media, but alignment media were identified which enable extraction of the total coupling constants.

6.4 Experimental

Synthesis of achiral polymers

The syntheses presented are generally inspired from the literature [4, 7, 16, 17]. All commercially bought monomers and crosslinkers (except *N,N*-methylenbisacrylamide which is a solid) were purified prior to the synthesis to remove the polymerization inhibitor by passing the neat liquid through a pipette filled with basic alumina.

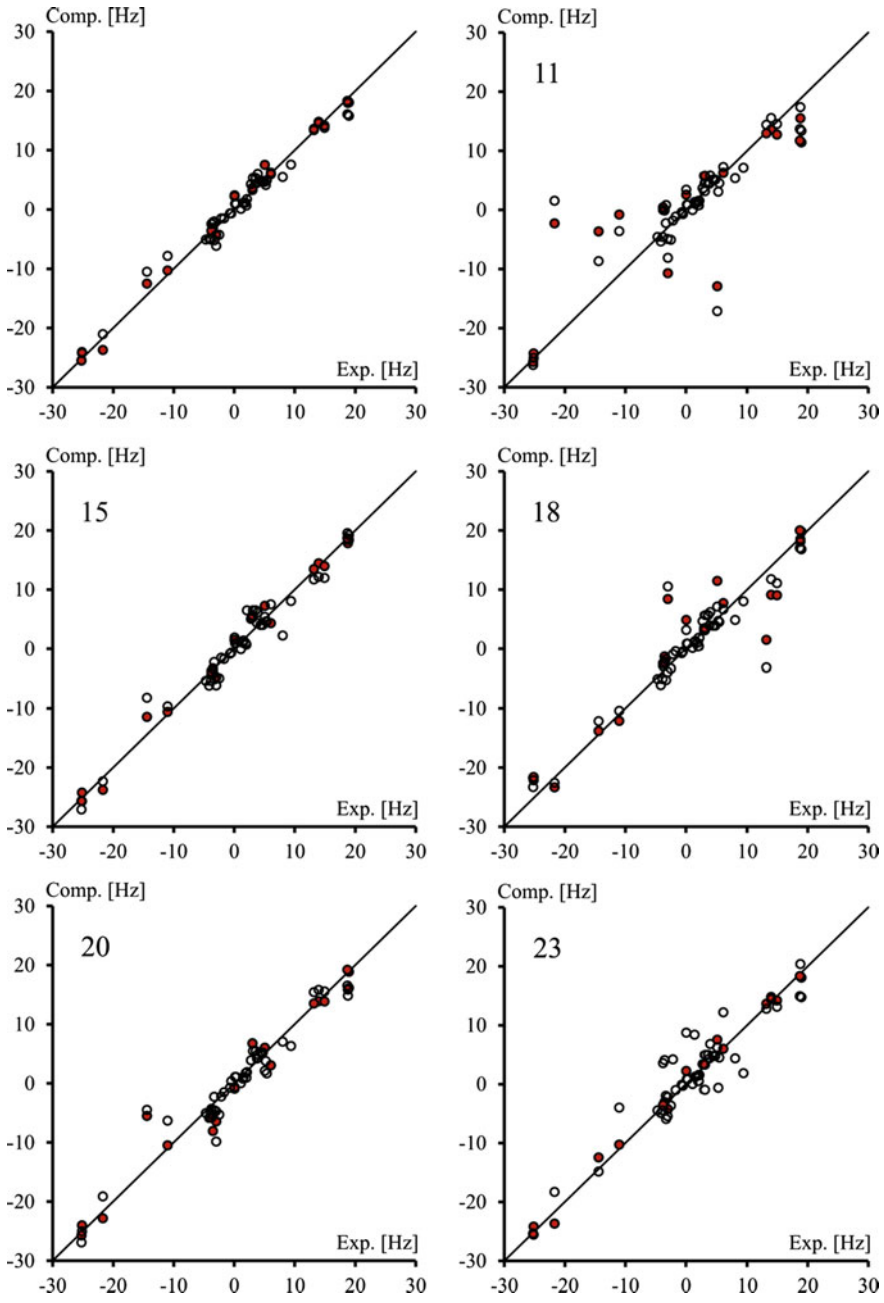
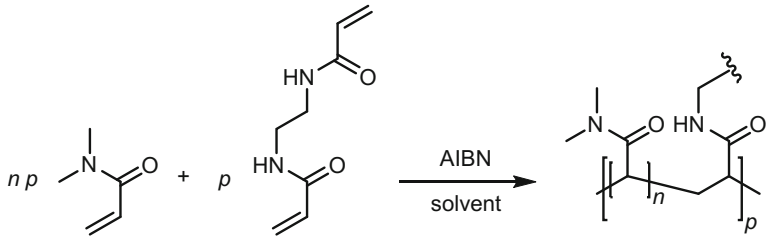


Fig. 6.5 Comparison of experimental and back-calculated RDCs of strychnine. Red dots: only $^{1}D_{CH}$ data. Open circles: $^{1}D_{CH}$ and $^{n}D_{HH}$ from S^{3} HMBC *homo*. Upper left is right stereochemistry, and the diastereotopic protons are switched in the rest as indicated in the plots. Anisotropy induced by PDMAA (Polymer 6.2). Tables with the data are found in Appendix A4

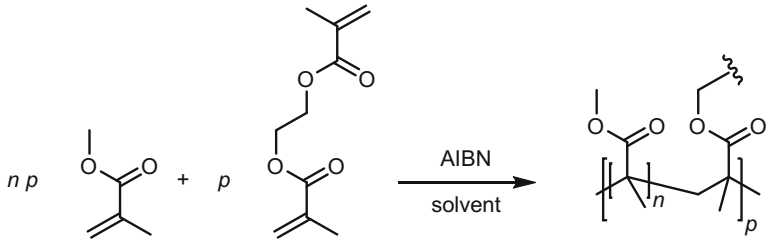
PDMAA



To 1 mL of *N*,*N*-dimethyl acrylamide (9.7 mmol) was added 1.6 mg (0.009 mmol) *N*,*N*-methylenbisacrylamide and 1.5 mg azobisisobutyronitrile (AIBN) (0.009 mmol). N_2 was bubbled through the solution for 5 min to remove O_2 and the solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and N_2 backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14–16 h) to polymerize.

A solvent ($CDCl_3$, DCM, acetone-d_6) may be added before transfer, and was usually used in 1:2 ratio of solvent: *N*,*N*-dimethyl acrylamide (*v:v*).

PMMA



The protocol of Gil et al. was generally followed [4].

10 mL MMA (61.0 mmol), was added 2 mL acetone-d_6 and 2.5 mg AIBN (0.015 mmol). 10 mL were taken out and mixed with 40 μL ethylene glycol dimethylacrylate (EGDMA) (0.42 mmol). The solution was transferred to 3 or 5 mm NMR tubes evacuated and back-filled with N_2 three times. The cycle was repeated 3 times, and the tubes were left at 50 °C overnight (approx. 14–16 h).

Swelling of polymers
5 mm sticks

The tubes were left for a couple of hours without lid after being cooled to room temperature. The polymers could usually be removed easily from the tubes without need of breakage. If stuck the tubes were carefully broken though. The polymers were then cut to appropriate length (2 cm) and inserted into a 5 mm NMR tube. The

solvent and analyte were added and the polymer was left to swell. Depending on the polymer the swelling time varied from approximately one week to several weeks. The degree of alignment was carefully monitored from the quadrupolar splitting of the deuterated solvent.

3 mm sticks

The tubes were carefully broken and left overnight to dry. This was especially beneficial for the more sticky chiral polymers. The glass was carefully removed from the polymers and the resulting polymer sticks were cut into the appropriate length (4 cm). The cut polymer stick was inserted into a 5 mm NMR tube and the analyte was dissolved in 0.3 mL solvent and added. A plug was quickly inserted to prevent vertical swelling. Samples should be left for at least 5 h ($CDCl_3$), 1 d (DMSO-d_6) or 2-3 d (CD_3OD) for swelling. To make sure that complete swelling was achieved, a minimum 1d ($CDCl_3$), 3 d (DMSO-d_6) or 5 d (CD_3OD) was always used.

In the table below the constituents in the synthesis of the relevant polymers are given.

Name	Analyte (c [mM])	Solvent	Monomer	Crosslinker (mol%)	Monomer:solvent in synthesis (v:v)	Field (MHz)
Polymer 6.1	IPC (553.6)	$CDCl_3$	PMMA	EGDMA (0.38)	4.5:1 acetone-d_6	800
Polymer 6.2	Strychnine (216.5)	$CDCl_3$	DMAA	MBAA (0.34)	–	800

NMR experiments

The CLIP-HSQC experiment was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. Isotropic spectra of IPC (145 mM) and strychnine (144 mM) were acquired in 5 mm tube (0.5 mL).

The S^3 HMBC *homo* was used for long-range homonuclear coupling constants as described in Chap. 4, using standard setup as described in Appendix A1.

Total coupling constants were extracted from aligned samples, while *J*-coupling constants were extracted from isotropic samples, in order to determine the RDCs, as described in Sect. 6.5.1.

Simulations

Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations, [18] using the program MacroModel version 10.8 [19, 20]. The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization [19, 21].

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations [22]. Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

References

1. J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, J. Magn. Reson. **138**, 334–342 (1999)
2. C.M. Thiele, A. Maliniak, B. Stevensson, J. Am. Chem. Soc. **131**, 12878–12879 (2009)
3. A. Marx, C. Thiele, Chem. Eur. J. **15**, 254–260 (2009)
4. C. Gayathri, N.V. Tsarevsky, R.R. Gil, Chem. Eur. J. **16**, 3622–3626 (2010)
5. J.C. Freudenberger, S. Knör, K. Kobzar, D. Heckmann, T. Paululat, H. Kessler, B. Luy, Angew. Chemie Int. Ed. **44**, 423–426 (2005)
6. P. Haberz, J. Farjon, C. Griesinger, Angew. Chemie Int. Ed. **44**, 427–429 (2005)
7. M. Schmidt, H. Sun, A. Leonov, C. Griesinger, U.M. Reinscheid, Magn. Reson. Chem. **50**, S38–S44 (2012)
8. T. Parella, J.F. Espinosa, Prog. Nucl. Magn. Reson. Spectrosc. **73**, 17–55 (2013)
9. B.L. Marquez, W.H. Gerwick, R.T. Williamson, Magn. Reson. Chem. **39**, 499–530 (2001)
10. F. Tian, P.J. Bolon, J.H. Prestegard, J. Am. Chem. Soc. **121**, 7712–7713 (1999)
11. J. Yan, F. Delaglio, A. Kaerner, A.D. Kline, H. Mo, M.J. Shapiro, T.A. Smitka, G.A. Stephenson, E.R. Zartler, J. Am. Chem. Soc. **126**, 5008–5017 (2004)
12. A. Möglich, M. Wenzler, F. Kramer, S.J. Glaser, E. Brunner, J. Biomol. NMR **23**, 211–219 (2002)
13. N.C. Meyer, A. Krupp, V. Schmidts, C.M. Thiele, M. Reggelin, Angew. Chemie Int. Ed. **51**, 8334–8338 (2012)
14. A. Marx, V. Schmidts, C.M. Thiele, Magn. Reson. Chem. **47**, 734–740 (2009)
15. A. Enthart, J.C. Freudenberger, J. Furrer, H. Kessler, B. Luy, J. Magn. Reson. **192**, 314–322 (2008)
16. B. Luy, K. Kobzar, S. Knör, J. Furrer, D. Heckmann, H. Kessler, J. Am. Chem. Soc. **127**, 6459–6465 (2005)
17. A. Krupp, M. Reggelin, Magn. Reson. Chem. **50**, S45–S52 (2012)
18. Maestro (Schrödinger, LLC, New York, 2015)
19. Schrödinger, (LLC, New York, NY, 2016)
20. Desmond (D. E. Shaw Research, New York, 2015)
21. G. Chang, W.C. Guida, W.C. Still, J. Am. Chem. Soc. **111**, 4379–4386 (1989)
22. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K. N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V. G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in *Gaussian 09, Revision B.01* (Gaussian Inc., Wallingford, CT, 2010)

Chapter 7
Chiral Alignment Media for Enantiodiscrimination

Nature is chiral—and most biological molecules are chiral structures [1, 2]. Usually one enantiomer of a given macromolecule or small molecule is universally present in an organism, leading to highly specific chiral environments. In line with the inherent chirality of nature, the chirality of target molecules in organic synthesis, as well as molecules isolated from natural sources, is of utmost importance. This is a consequence of the compounds often being interesting due to interactions with a natural target [1, 2]. The practical importance may be exemplified by a drug hoping to reach the US market. To be qualified for approval, the FDA demands that the compound is enantiomeric pure, or that both enantiomers are evaluated [3]. This is due to often very different pharmacological properties of enantiomers caused by the chiral environment of the human body [1].

While chiral centers may usually be solved from a single known stereocenter, and diastereomers thus may be separated, enantiomeric structures are much harder to resolve. Similar physical properties of the enantiomers result in identical observables in the achiral environments where small molecules are usually investigated, be that utilizing MS, IR, UV or NMR methods. One of the major selling points for RDCs in small molecular NMR, besides providing an orthogonal method to e.g. NOE connectivities and *J*-coupling constants in obtaining 3D structural information, is a promise of differentiation of enantiomers. This is arguably the most challenging stereochemical problem to solve in organic as well as natural product chemistry, and thus a premium goal for NMR spectroscopy.

7.1 Methods of Enantiodiscrimination

Due to the important functional differences of enantiomers, the differentiation of these has been a major research topic for years. RDCs are a relatively new approach, and other methods will thus briefly be mentioned. While NMR spectroscopy is utilized in some of the methods, many utilize different analytical tools.

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_7

The probably most utilized method of discerning enantiomers is by optical rotation, where it is utilized that the enantiomers will rotate plane-polarized light in opposite directions [4]. The analysis is carried out by shining single-wavelength, plane-polarized light through a solution of the analyte and the rotation is measured; a negative value means a rotation to the left, and a positive to the right. The specific rotation [α] is given by Eq. (7.1), at 20 °C and using a sodium lamp of wavelength 589 nm.

$$[\alpha]_D^{20} = \frac{\alpha}{c \cdot l} \tag{7.1}$$

where α is the observed rotation, c is the concentration of the analyte and l is the path length of the light. The rotation is dependent on chirality as well as other structural features and a reference of known chirality is needed to determine the absolute stereochemistry [4].

Another widespread method to distinguish enantiomer is by chemical derivatization, where a specific functional group is derivatized using a chiral building block of known absolute stereochemistry. The derivatization changes the enantiomers to diastereomers, which may be separated by e.g. standard HPLC systems, an may lead to different resonances in NMR spectra. A multitude of reagents are available dependent on the nature of the compound which is investigated [5–9]. Some of the more well-known include Mosher's reagent for, primarily, derivatization of secondary alcohols and Marfey's reagent for amino acids [10, 11]. Using Marfey's reagent as an example, the diastereomers are compared to diastereomers of known chirality, synthesized from pure amino acids, as illustrated in Fig. 7.1. The subsequent separation on a HPLC column obtained by Marfey's reagent is increased by steric differences in the structures resulting in stronger intramolecular hydrogen bonds of the L-isomer [10].

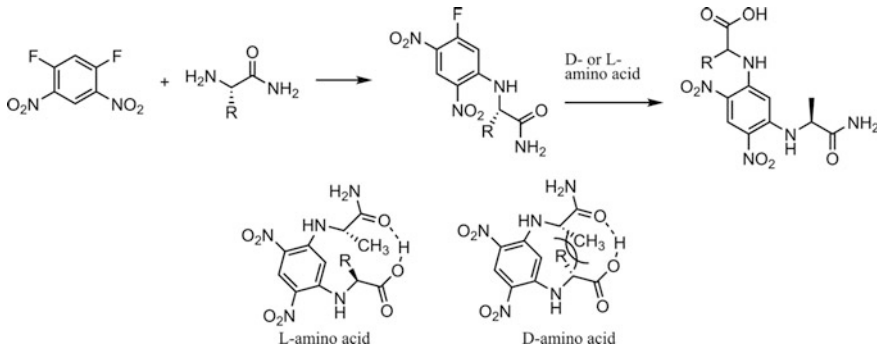


Fig. 7.1 Top: Marfey's reagent, made from 1, 5-Difluoro-2, 4-dinitrobenzene and L-alanine-NH_2 which is reacted with an amino acid. Bottom: The difference in the spatial properties of the complex when reacted with L- or D-amino acids

Of methods to differentiate enantiomers by NMR other than RDCs, paramagnetic lanthanide chemical shift reagents and chiral solvating agents should be mentioned [8, 9]. The lanthanide reagents are complexes of enantiomerically pure organic structures, e.g. based on camphor derivatives, coordinated to a lanthanide ions (Eu, Pr, Yb are mostly used). Coordination to the lanthanide complex will result in a change in chemical shifts of the nuclei in the analyte dependent on the distance of said nuclei to the lanthanide, and enantiomers will interact differently with the chiral complex, leading to enantiodiscrimination [8]. Diamagnetic chiral solvating agents work in a similar way, though the induced differences in chemical shifts are usually smaller. Here a chiral additive is added to the solvent, which forms complexes with the analyte and will result in a change in chemical shifts dependent on the enantiomer, due to differences in interactions. The chiral solvating agents rely heavily on the magnetic anisotropy induced by aromatic groups in the structure of the additives [8, 9].

Also chiral HPLC columns, single crystal X-ray crystallography and vibrational or electronic circular dichroism may be used to discern enantiomers [3].

7.2 Synthesis of Media

To widen the pool of enantiodiscriminating alignment media, the focus was decided to be on stretched polymers, since quite good enantiodiscriminative alignment media are available in the LC pool, and due to the inherent problems of strong alignment of LCs. It was thought useful if the enantiodiscrimination of more general compounds than charged amines would be possible utilizing stretched polymers, and it was proposed to accomplish this by combining the conclusions reported for the polyacetylenes and chiral APhES polymers [12, 13].

The chiral media synthesized were all based on a simple strategy; a chiral building block was attached to a double bond and polymerized with a cross-linker, as illustrated in Fig. 7.2. Thus the chiral center is situated in the side-chain of the polymer, while the backbone must be considered a racemic mixture of stereocenters. It is realized that a set chiral center on a side-chain may influence the chirality

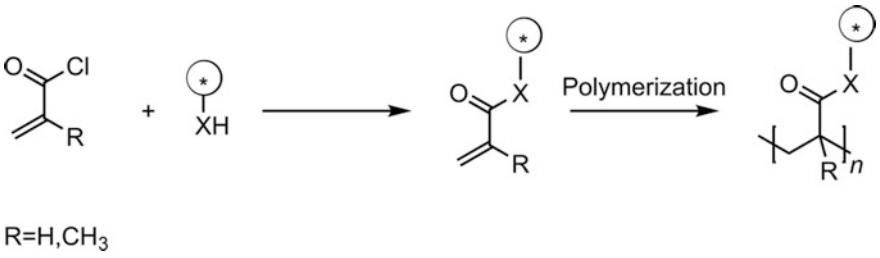


Fig. 7.2 Reaction scheme of the polymerization of chiral monomers (cross-linker not shown)

of the stereocenter set during polymerization, but the tacticity of the polymers was not investigated further [14–18]. The differentiation of enantiomers is proposed to be based on differences in interaction to the chiral centers of the static polymer, leading to a different average orientation in space between enantiomers in the aligned samples.

The attachment of the double bond to the chiral molecule was achieved by reaction of a secondary alcohol or amine and an acid chloride. This was followed by polymerization in presence of a cross-linker and possibly a co-monomer, using azobisisobutyronitrile (AIBN) as the initiator, outlined in Fig. 7.3.

The chiral molecules used are depicted in Fig. 7.4. The molecules were chosen to resemble previously successfully employed monomers from LC and SP alike [12, 13, 19]. It was important that the dispersion of resonances from the polymer was as small as possible, as residual signals were present in the spectra upon alignment. This limited the amount of possible chiral building blocks to simple structures, or structures with many overlapping shifts, e.g. from long homogenous chains. The importance of this is illustrated in Fig. 7.8 p. 160, where a spectrum of an aligned sample of strychnine in $CDCl_3$ is shown. The chosen monomers displayed resonances in two specific regions of NMR spectra: one in the aromatic region and one in the aliphatic region.

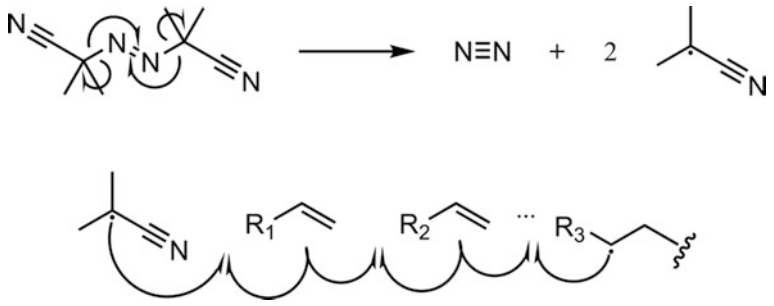


Fig. 7.3 Polymerization reaction using AIBN as an initiator, followed by propagation prolonging the chain until the polymerization is terminated from reaction to a second radical. The initial reaction is induced by heating

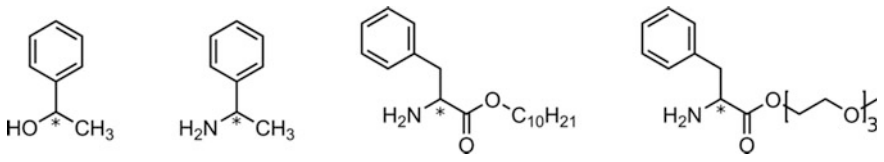


Fig. 7.4 The structures of the chiral building blocks used in the work of this thesis. From left: 1-phenyl ethanol (PhEtO), 1-phenyl ethylamine (PhEtN), $C_{10}H_{21}$-phenylalanine (C_{10}-Phe) and methyl-trietyleneglycol-phenylalanine (TEG-Phe)

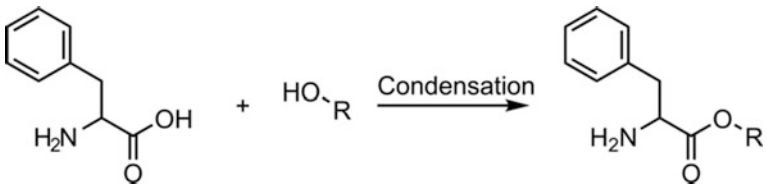


Fig. 7.5 Condensation reaction of phenylalanine under acidic conditions

The first two chiral building blocks used, (*S*)-1-phenyl ethanol (PhEtO) and (*R*)-1-phenyl ethylamine (PhEtN), were readily available as pure enantiomers, while the remaining two were based on *L*- or *D*-phenylalanine and dubbed $C_{10}H_{21}$-phenylalanine (C_{10}-Phe) and methyl-trietyleneglycol-phenylalanine (TEG-Phe) in the following, synthesized as indicated in Fig. 7.5.

For more on the synthetic condition see the experimental section.

7.3 Results

The swelling properties of the resulting polymers were evaluated similarly to the achiral polymers described in Sect. 6.1. In all cases, in line with available literature and previous findings for the PDMAA polymer, the radial swelling of a thin polymer was superior to vertical swelling of a short polymer, see Fig. 5.6 for an illustration [13, 20]. This is in part contributed to the fact that the thicker polymer sticks used for vertical swelling formed a plug when solvent was added, which hindered further vertical swelling. Thus only very small RDCs were observed. The nature of the study, comparing enantiomers, also required a high degree of reproducibility, where radial swelling was determined to be inherently superior due to more control of the swelling. Experiments showed that it was much easier to cut the right length of a polymer stick and swell this under identical restriction, than hoping that the polymers would stop swelling at identical lengths. One could of course restrict vertical swelling as well, had plugs not been formed upon the addition of solvent. Another argument in favor of using the radial swelling approach was the time needed for swelling. While vertical swelling may take weeks to months for certain polymers [20, 21], radial swelling was visibly finished and experimentally stable, in regards to the extracted RDCs, in a matter of 5–10 h in chloroform and 1–2 days in DMSO. No to minor [2] H-splitting of the solvent was observed for the polymers in all solvents, in contrast to the native achiral polymers, which for the most part was PDMAA (see Sect. 6.1).

Interfering differences in observed RDCs were determined for identical compounds aligned in identical polymers, which were contributed to random errors as well as small differences in the length of the utilized polymer sticks. This was corrected by choosing a reference RDC across all samples and comparing the relative differences in the observed RDCs, and led to better reproducibility. As the difference

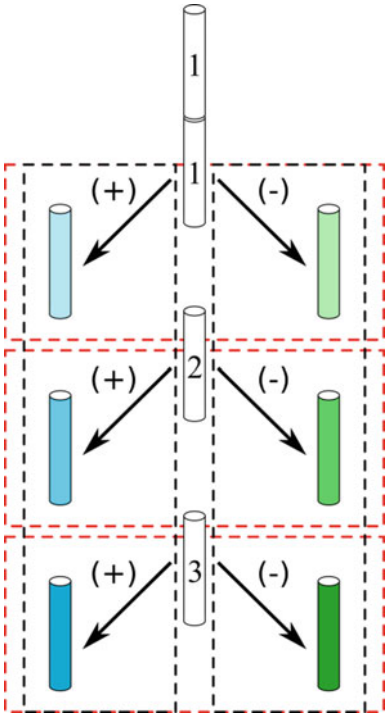


Fig. 7.6 The evaluation of enantiodiscrimination of the alignment media was made by comparing enantiomers, (+) or (−), aligned in polymers from a single polymer stick (red dotted boxes). This was repeated using multiple polymer sticks (2,3…), so that the variation for identical compounds between polymers was also evaluated (black dotted box). This gives rise to two comparisons which are made and depicted in the following sections: 1. The averages and standard deviations of RDCs from each black box are compared. 2. The RDCs of the red boxes are compared as averages and standard deviations of $RDC_{(+)}$—$RDC_{(-)}$ for each red box. This is equal to the difference in alignment, disregarding any possible differences in the alignment between polymer sticks

in RDCs between enantiomers is proposed to be their relative orientation compared to the magnetic field, and scaling all RDCs is equal to altering the alignment strength D_a, using the relative RDCs will lead to the right conclusion, while eliminating errors from e.g. different polymer lengths. The reference RDC was chosen as a large RDC with a low variance between all data sets, regardless of stereochemistry. The comparison was made utilizing a setup as explained in Fig. 7.6.

7.4 PhEtO, PhEtN and C_{10}-Phe Polymers

It was realized that no enantiodiscrimination could be attributed to the polymers made from (*S*)-1-phenyl ethanol, (*R*)-1-phenyl ethylamine and *L*-$C_{10}H_{21}$-phenylalanine or at least the enantiodiscrimination was below the experimental error and

differences between identical compounds. Thus the results will only be shortly presented.

It was possible to synthesize a polymer made entirely of (*R*)-PhEtO, as the monomer was a liquid and solubility was not an issue. The differentiation of enantiomers was not observed though, as illustrated using the enantiomers of IPC in Fig. 7.7. The variance between identical compounds was 0.72 and 0.47 Hz between enantiomers, and it was concluded that enantiodiscrimination was not achieved.

The same trend was observed for the (*R*)-PhEtN polymer, and since it was structurally alike to the (*R*)-PhEtO polymer, while having some solubility difficulties during synthesis, further evaluation was not made. The (*R*)-PhEtN polymer swelled in $CDCl_3$, DMSO-d_6 and methanol-d_4, giving rise to quite different size and sign of RDCs between the apolar and polar solvents.

The (*L*)-C_{10}-Phe polymer may differentiate enantiomers, but reproducibility was difficult due to poor solubility of the monomer during synthesis, which resulted in brittle polymers and ultimately large differences in the observed RDCs. A similar LC to the (*L*)-C_{10}-Phe polymer has been reported to work well, even at temperatures where no backbone order was supposedly present, and it was surprising that such a low, if any, differentiation was observed [12]. The low solubility of the monomer may be the problem, as the ratio of chiral to achiral monomer might have been too low to achieve observable differentiation.

The solubility problem of the (*L*)-C_{10}-Phe polymer was alleviated by using methyl-tri-ethylene glycol as the long ester chain and more of this monomer could be incorporated into the polymer. TEG-phenylalanine based polymers showed minor enantiodiscriminative properties and the results are included below.

7.5 The TEG-Phe Polymer

The (*L*)-TEG-Phe polymer exhibited much better solubility properties compared to the (*L*)-C_{10}-Phe polymer, while retaining approximately the same size of the ester group (though the structure of the chain in solution may vary). This polymer

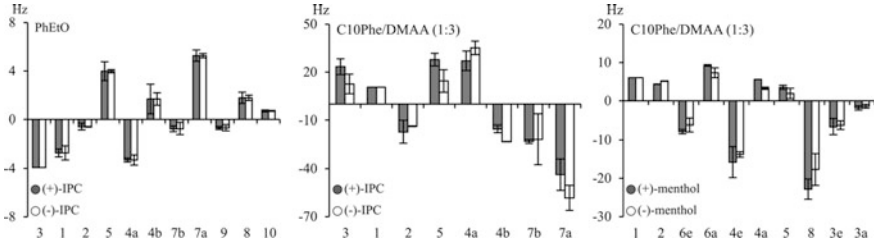


Fig. 7.7 Comparison of RDCs from enantiomers in different chiral media, as given in the plots. Error bars indicate that a mean is used with standard deviation of three datasets for each enantiomer

became the focus of the study, and a more in depth analysis of the alignment properties is included. The polymer swelled in $CDCl_3$, DMSO-d_6 and methanol-d_4 giving rise to RDCs in the same general magnitude, but only $CDCl_3$ and DMSO-d_6 were investigated in depth. Similar to the PhEtN polymer, a big difference in extracted relative RDCs was observed between apolar and polar solvents, leading to a proposed difference in alignment, *vide infra*.

The polymer displayed residual resonances grouped in an aromatic ($\sim$7–7.5 ppm/120–125 ppm) and a TEG part ($\sim$3–4.5 ppm/60–75 ppm) with the latter leading to the most intense signals, illustrated in Fig. 7.8. The backbone-, α- and β-resonances were much less pronounced in the spectra, possibly due to less flexibility leading to a shorter T_2. The observed analyte resonances were broadened, which may be problematic, due to an increased uncertainty in the extracted coupling constants in the F2 dimension. The determination of the coupling constants was performed by displacement of 1D slices, where the peaks are overlaid to access the values, and the broad peaks should thus not be that problematic as long as they are symmetric.

Alignment and differentiation of the enantiomers of the small molecules IPC and menthol were pursued, and a single enantiomer of strychnine was also investigated, using polymers of opposite side-chain chirality.

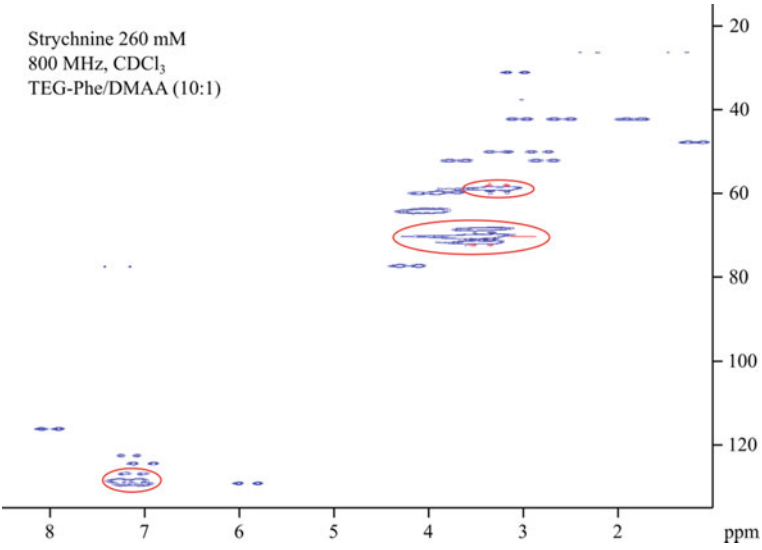


Fig. 7.8 Spectrum of strychnine in $CDCl_3$ aligned in 3 mm TEG-Phe/DMAA polymer stick. Residual polymer signals are highlighted. The only indeterminable strychnine resonance due to spectral overlap is C3–H3 in the aromatic region. The polymer resonances are well contained in small regions of the spectra. (800 MHz, 16 scans, rd = 1.5 s, 4 k × 256 zero-filled to 16 k × 1 k)

7.5.1 Menthol

Menthol is a molecule with a rigid chair conformation due to three substituents in equatorial positions. A problem in using menthol in RDC comparisons, using only $^1D_{CH}$-coupling constants, is that only four non-parallel internuclear vectors are present, unless the isopropyl orientation is considered known, see Fig. 7.9. Consequently, SVD fittings were only illustratively performed from the data. For more details on SVD, the reader is referred to Sect. 8.1.

The general trends of the extracted RDCs could be investigated though, see Fig. 7.10, and it was observed that the equatorial protons of the enantiomers had different signs in DMSO. This is a clear sign of enantiodiscrimination as the angle to the magnetic field must differ, independent of alignment strength. A similar trend of differentiation of the equatorial protons, although less pronounced, was determined from alignment in $CDCl_3$ (Fig. 7.10). Though the structure may not be fixed in space by SVD, the data suggested that the structure was aligned differently between the two solvents, from the sign of the RDCs of the axial protons. Reproducing the RDC results was more difficult in chloroform compared to DMSO, which may be contributed to rapid swelling in the solvent. This correlates well to an observation of the RDCs being markedly different if the insert in the tube had visibly moved. All samples with a displaced insert were discarded, but small displacement, unnoticeable to the human eye cannot be dismissed.

The data was fitted using SVD, though the fit is underdetermined and the structure is placed due to minor difference in bond lengths and angles for parallel vectors in the chair-conformation. Thus experimental errors will have a huge effect. The difference in the experimental RDCs, defined as the generalized cosine β angle (GCB), between identical stereoisomers was 0.996 ± 0.001 or $4.8 \pm 0.8°$ and 0.943 ± 0.007 or $19.3 \pm 1.3°$ between enantiomers, illustrated in Fig. 7.11 by 3D structures of (+)-menthol turned into the alignment frames. The definition of GCB is given in Sect. 5.1.4 p. 137.

In conclusion, menthol was differentiated by alignment in the (*L*)-TEG-Phe polymer, especially apparent in DMSO, while the exact difference in alignment frame was not accurately determined.

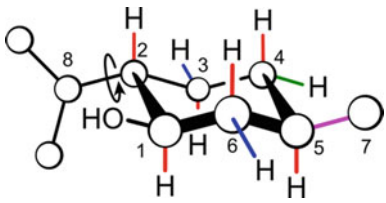


Fig. 7.9 The structure of menthol. Equally colored CH-bonds indicate that the bonds are parallel

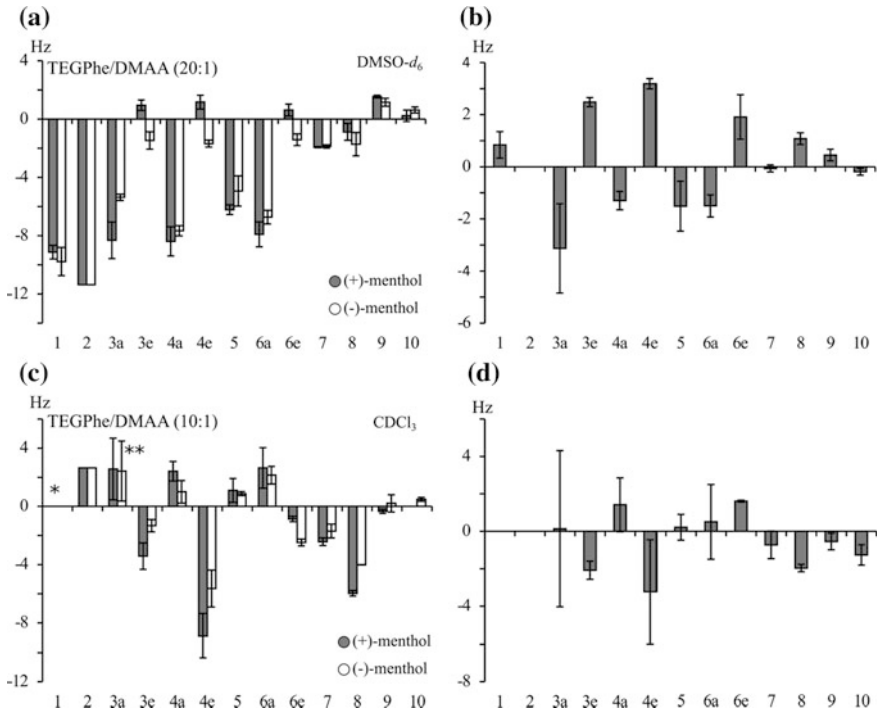


Fig. 7.10 Comparison of RDCs from enantiomers of menthol in (*L*)-TEG-Phe/DMAA. Solvent is DMSO-d_6 in (**a**) and (**b**), and $CDCl_3$ in (**c**) and (**d**). In DMSO three datasets were used for (+)-menthol and two for (−)- menthol. In $CDCl_3$ two datasets were used for each. (**b**) and (**d**) display values of (−)-menthol subtracted from (+)-menthol in rods cut from identical polymer sticks. A mean is displayed with standard deviation of three datasets for each enantiomer. *Omitted due to overlap with polymer resonances. **Overlaps with CH-7

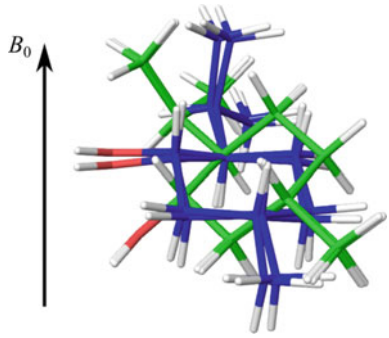


Fig. 7.11 Comparison of the alignment from RDCs of (+)- and (−)-menthol in DMSO-d_6 aligned in (*L*)-TEG-Phe/DMAA. Both datasets were applied to the structure of (+)-menthol, green are (−)-menthol and blue (+)-menthol datasets

7.5.2 IPC

The small molecule IPC (Fig. 7.12) was also investigated in DMSO and chloroform. Similar comments to those of menthol on the reproducibility between solvents may be made, and only the data from DMSO are presented.

The RDCs of aligned enantiomers in DMSO are compared in Fig. 7.13, with the dataset including three repetition for each enantiomer. The differences in observed RDCs were small, and again the differences in the used polymer sticks were eliminated by subtracting the RDCs of (−)-IPC from those of (+)-IPC, aligned in sticks cut from the same polymer. A pattern was identified, though the deviations were substantial compared to the observed differences. The GCB between enantiomers was determined to be 0.987 ± 0.005, compared to 0.997 ± 0.002 for identical compounds. This is equal to a difference in angle between alignment frames of 8.9 ± 2.0° between enantiomers compared to 4.4 ± 1.6° for identical compounds. It is thus not a significant differentiation, and it may be argued that the differences are too low to support enantiodiscrimination of IPC due to comparable differences between enantiomers and the reproducibility of the data. Compared to the GCB of PBLG, which has been reported as 0.991, the presented polymer seems

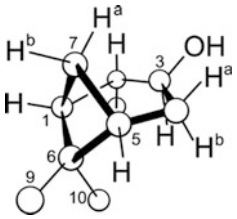


Fig. 7.12 Structure of (+)-IPC

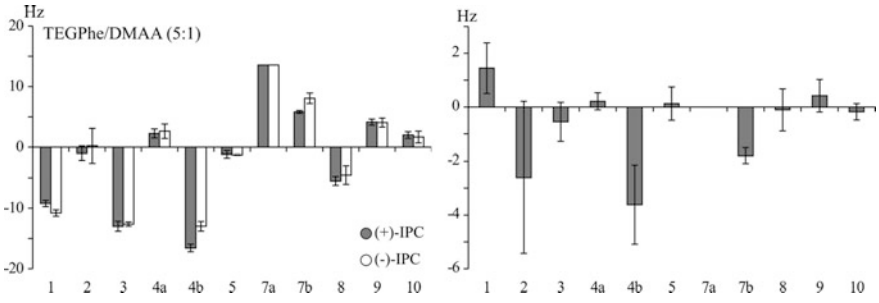


Fig. 7.13 Comparison of RDCs from enantiomers of IPC in (*L*)-TEG-Phe/DMAA (left). Solvent is DMSO-d_6. Right: Values of (−)-IPC subtracted from (+)-IPC in rods cut from identical polymer sticks. A mean is used with standard deviation of three datasets for each enantiomer

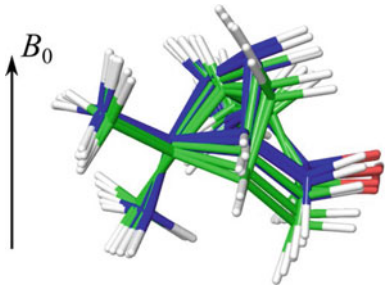


Fig. 7.14 Comparison of the alignment from RDCs of (+)- and (−)-IPC in $CDCl_3$ aligned in (*L*)-TEG-Phe/DMAA. Both datasets were applied to the structure of (+)-IPC, green are (−)-IPC and blue (+)-IPC datasets. The enantiomers are grouped in two, slightly overlapping groups

to be a slightly better enantiodiscriminator, but exhibits less reproducibility, limiting the credibility [12, 19]. The alignment frames of the enantiomers are compared in Fig. 7.14.

7.5.3 *Enantiomeric Polymers*

It is possible to synthesize monomers starting from both enantiomers of phenylalanine and it was thus possible to check whether e.g. strychnine, where only one enantiomer is readily available, could be differentiated by the polymers. Comparing RDCs from the *D*-polymer for (±)-menthol to data of the enantiomers from the *L*-polymer, indeed opposite RDCs were obtained, shown in Fig. 7.15, where RDCs of (−)-menthol in *D*-TEG-Phe are similar to those of (+)-menthol in *L*-TEG-Phe and vice versa. For the IPC enantiomers the result was less convincing, as seen in Table 7.1. An increase in the GCB of IPC enantiomers in different media was observed compared to identical compounds in identical media, especially for

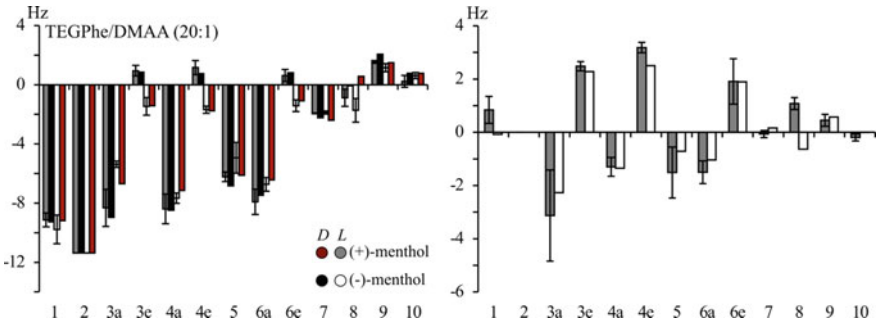


Fig. 7.15 Left: Comparison of RDCs from enantiomers of menthol in TEG-Phe/DMAA. Solvent is DMSO-d_6. Grey (+) and white (−) are the enantiomers in *L*-Phe based polymers, while black (−) and red (+) are enantiomers in *D*-Phe based polymers. Right: Values of (−)-menthol subtracted from (+)-menthol in rods cut from identical polymer sticks. Grey is *L*-Phe and white is *D*-Phe based polymer. A single dataset for the enantiomers in D-Phe based polymers was used

Table 7.1 Comparison of the alignment tensors of enantiomers of IPC from *L* and *D*-variant of the TEG-Phe based alignment media

		L-Phe		*D*-Phe	
		(+)-IPC	(−)-IPC	(+)-IPC	(−)-IPC
L-Phe	(+)-IPC	*5.0 ± 1.7°*	**8.9 ± 2.0°**	**9.6 ± 3.1°**	*8.2 ± 2.0°*
	(−)-IPC		*1.7 ± 1.5°*	*4.1 ± 1.3°*	**9.9 ± 2.2°**
D-Phe	(+)-IPC			6.4°	**9.7 ± 3.0°**
	(−)-IPC				4.8°

The comparison is performed by the GCB. Comparisons which should lead to equal alignment are marked in italic, and different alignments in bold, if enantiodiscrimination is achieved. Dataset of *L*-polymers includes three sets of RDCs for each enantiomer, for the *D*-polymer two are included

(+)-IPC in *L*-TEG-Phe and (−)IPC in *D*-TEG-Phe, where the difference nears the difference between the enantiomers in both polymers. While this translates to a difference in orientation between the enantiomeric polymers, identical trends in RDCs were observed. Again it would seem that the enantiomers of IPC were not significantly differentiated. This also shows that conclusions based on RDCs obtained between enantiomeric polymers needed to be made carefully.

7.5.4 Strychnine

As the enantiomeric polymers aligned single enantiomers differently, at least in the case of menthol, a theoretical differentiation of enantiomer of strychnine was approached, similar to the work of Schmidt et al. [13]. Not all one-bond vectors were used for comparison, as the resonance of C3–H3 was overlapped with polymer resonances in the CLIP-HSQC spectra and the RDCs could not be determined. The resonances of C-15 were weak and slightly distorted compared to the other resonances and were omitted (the extracted RDCs were of equal size between all samples). The resonances of C-23 were not separated, in line with experiences from other media (Sect. 6.2). As a result 15 RDCs were extracted, and a difference was observed between the polymers based on *L*- and *D*-phenyl alanine, as illustrated in Fig. 7.16.

The data was based on a triple study, using three different *L*- and *D*-polymer sticks. While it was not possible to compare alignment from equal polymer sticks, as an identical enantiomer of strychnine was used throughout, the alignment was compared based on the succession of obtain data, i.e. polymer L_1 is compared to D_1, L_2 to D_2 and L_3 to D_3, in Fig. 7.16. The used polymers sticks were synthesized from the same bulk reaction mixture, but polymerized in different tubes, as to take differences in polymerization into account.

Rather big differences in RDCs between strychnine aligned in the two media were observed, and enantiodiscrimination was, cautiously, considered achieved. Examples of 1D slices through the CLIP-HSQC are found in Fig. 7.17.

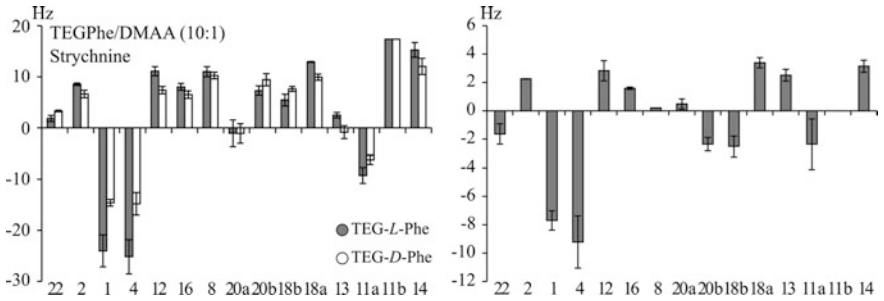


Fig. 7.16 Left: Comparison of RDCs from strychnine in (*L*)-TEG-Phe/DMAA (grey) and (*D*)-TEG-Phe/DMAA (white). Solvent is $CDCl_3$. Right: Values obtained by subtracting polymer D_1 from L_1, D_2 from L_2 and D_3 to L_3. A mean is used with standard deviation of three datasets for each enantiomeric polymer

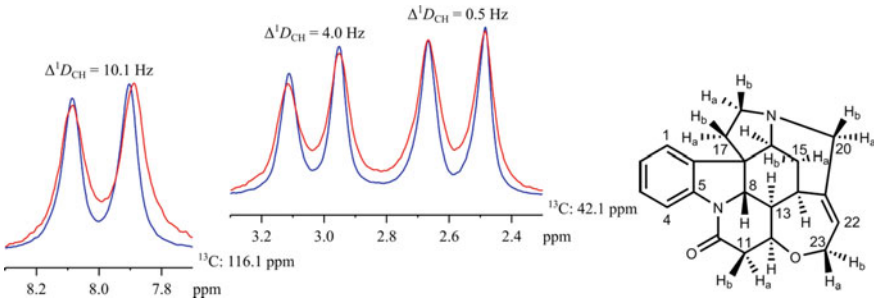


Fig. 7.17 1D slices from a CLIP-HSQC of strychnine in (*L*)-TEG-Phe/DMAA (blue) and (*D*)-TEG-Phe/DMAA (red) and the differences in the RDCs between the polymers. The total coupling constants were determined by displacement of peaks and comparison to extracted *J*-coupling constants (isotropic spectra not shown)

However, a problem was identified. While the SVD correlations of the experimental to back-calculated data for IPC enantiomers were generally good, with *Q*-factors between 0.085 and 0.210 (average 0.135), the correlation of the data to a strychnine 3D structure was bad, with *Q*-factors between 0.247 and 0.374 (average 0.314). It was investigated whether the problem could arise from the strychnine data being acquired at 800 MHz, compared to 400 MHz for IPC and menthol. When comparing datasets from 400 MHz to those of 800 MHz, the values varied slightly with an RMSD of 2.0 Hz, but the resulting correlation between experimental data and 3D structure was not improved compared to that of the 800 MHz data. Multiple resonances were less resolved at 400 MHz, the reason for acquiring the dataset at 800 MHz initially, and if these RDCs are omitted in the comparison the RMSD is 1.1 Hz between the two field strengths. The cause of the large difference between experimental and back-calculated data was not identified, and the reproducibility contradicts a conclusion of the difference being a product of erroneous extraction of

RDCs. Also known parallel vectors (e.g. C1–H1 and C4–H4) have equal RDCs, which initially indicated that experimental RDCs would fit back-calculated values.

When comparing alignment tensors, which will be less accurate due to the bad correlation between experimental RDCs and the SVD calculations, a difference between the two polymers was identified, see Fig. 7.18 for an illustration. The alignment frames from RDCs differ comparable to e.g. IPC in DMSO, with a GCB of 0.990 ± 0.006 or 7.4 ± 2.7° between identical media. Between different media the GCB was calculated to be much lower (resulting in a larger angle): 0.940 ± 0.037 or 19.0 ± 6.2°.

The alignment is markedly different compared to the PDMAA polymer, the pure polymer of the co-polymer used with TEG-Phe, as seen in Fig. 7.19. This may not be surprising due to the low amount of DMAA used in polymerization (10 mol% for strychnine samples). Back-calculation of RDCs in PDMAA placed the structure

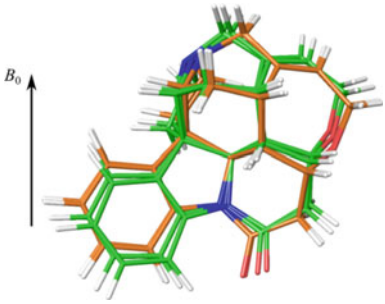


Fig. 7.18 Comparison of the alignment from RDCs of strychnine in (*L*)-TEG-Phe/DMAA (green) and (*D*)-TEG-Phe/DMAA (orange)

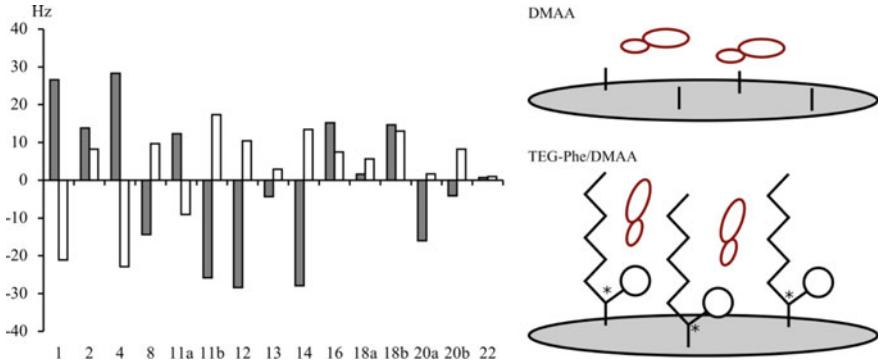


Fig. 7.19 Left: Example of RDCs from strychnine aligned in PDMAA (grey) and (*L*)-TEG-Phe/DMAA (white). Right: A possible scenario of alignment which explains the differences in RDCs. Red figures are analytes, the grey surface is an near-infinite polymer surface and the polymer side-chains and alignment of analytes are indicated

with the largest surface horizontal compared to the magnetic field, while the TEG-Phe polymer placed the largest surface parallel to B_0, perhaps best envisioned by the sign of the RDCs of C1–H1 and C4–H4 in Fig. 7.19. Since identical experimental procedures were followed, restricted swelling of 3 mm rods, this could translate to the TEG-chains extending from the surface with the analyte situated between the chains. The situations are illustrated by a cartoon in Fig. 7.19.

When stretched in DMSO, the polymer aligned solutes, such as IPC and menthol, perpendicular to the results in chloroform and more alike to the DMAA polymer (which also swells in DMSO).

7.6 Perspectives

For RDCs to impact the small molecule community in regard to enantiodiscrimination a couple of challenges need to be addressed, none of which are solved by the polymers introduced in this thesis. The largest problem is the need for both enantiomers of an analyte in comparison, due to a low difference in alignment between enantiomers. This is a problem in determining an unknown absolute stereochemistry of e.g. a natural product. For this feat to be realized, RDCs need to be more integrated with computational chemistry. The possibility to couple computational chemistry to RDCs and determine the absolute stereochemistry has been disputed, and it is acknowledged that the technology is not yet available to achieve this [22].

If the difference in alignment is large between enantiomers, as for e.g. shown for IPC aligned in the acetylenes [12, 19], and with further development of the methodology of Luy et al. where the alignment was determined including a polymer strand in simulation of strychnine, the possibilities are enticing [23]. If simulations could capture the differences in interactions of enantiomers to a chiral polymer strand, it should be possible to determine the absolute stereochemistry from a single enantiomer. This would require a difference in theoretical GCB of more than the 10–20° determined in this study, to get useful and trustworthy results and to negate errors from simulations and determined RDCs.

It would be interesting to investigate whether the TEG-Phe group, which showed better properties in stretched polymer, would retain those properties if utilized in LCs. This might be in LCs such as the acetylenes [12, 19], where an LC that aligns molecules in both chloroform and DMSO might be obtained. A proposed synthesis is outlined in Fig. 7.20.

Also other LCs might be of interest. Griesinger et al. recently published a LC made of graphene oxide with attached polymer brushes made of trifluoroethyl methacrylate, which displayed very interesting properties such as very narrow line widths and effectively no residual polymer signals [24]. The utilized solvent is DMSO, and due to the swelling properties, and thus solubility, of the TEG-Phe polymer in this solvent, it could easily be envisioned as an enticing prospect of adding chirality to the achiral graphene sheets and possibly obtain enantiodiscrimination.

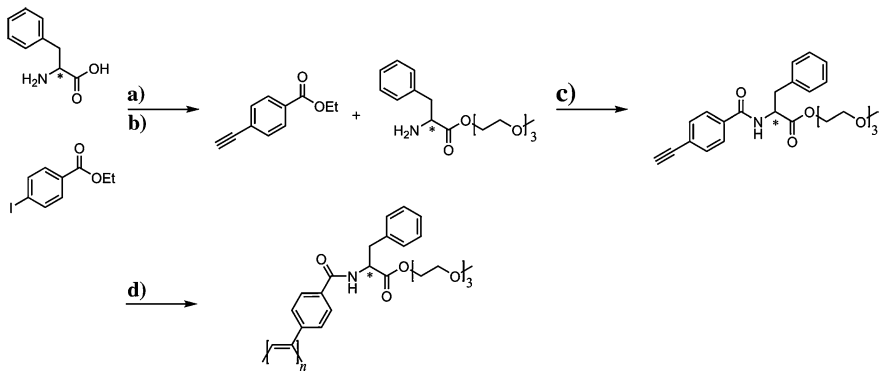


Fig. 7.20 Possible synthetic route to obtain an acetylene-based LC that might work in DMSO and $CDCl_3$, based on the stretched polymer work presented in this thesis. **a)** see experimental section. **b)–d)** see work of Krupp and Reggelin [12]

Based on the assumption that the (*L*)-C_{10}-Phe polymer failed to discriminate enantiomers due to a low amount of the chiral monomer incorporated into the polymers, it was not consistently tried to synthesize polymer with a molar fraction of TEG-Phe below 80%. In retrospect, it might be that a lower degree of the larger chiral monomer would lead to better access to the chiral centers, and thus better differentiation. An optimal molar fraction of chiral monomer for differentiation may thus be lower than the ones tested, which should be investigated further.

7.7 Conclusions

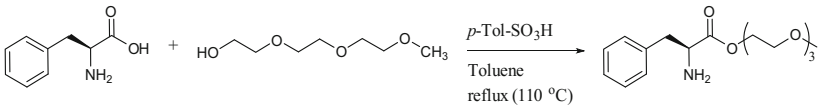
In conclusion a polymer was identified that possesses minor enantiodiscriminating properties comparable to PBLG in chloroform and DMSO. In principle a similar differentiation should be possible in methanol as the polymers align compounds in this solvent at an equal magnitude of alignment strength as DMSO, but this was not attempted. The line width might be problematic and further studies are needed to verify the differentiation. While being minor, the differentiation of enantiomers shows promise and this is the first uncharged radical-initiated polymer that has been shown to possess enantiodiscriminative properties. As for all alignment media, it is still a requirement to have each enantiomer as a standard for comparison, which renders the technology difficult and slow to use compared to e.g. optical rotation, which will result in similar, and possibly more significant, information regarding absolute chemistry.

7.8 Experimental

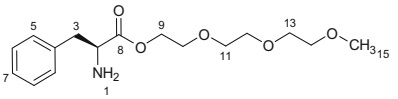
Synthesis of chiral polymers

The chemicals used were commercially available and the solvents used were all HPLC grade. Dry solvents were obtained from an in-house Pure Solv. All flasks were dried over a Bunsen flame under vacuum before any reactant or solvent was added. All flasks (or NMR tubes) were equipped with a rubber septum, and chemicals were transported by syringe. All commercially bought monomers and crosslinkers (except *N,N*-methylenbisacrylamide which is a solid) were purified prior to the synthesis to remove the polymerization inhibitor by passing the neat liquid through a pipette filled with basic alumina.

Preparation of (triethylene glycol methyl ether)-L-phenylalaninate



10 g L-phenylalanine (61 mmol) was suspended in 250 mL toluene, followed by 10.4 mL triethylene glycol methyl ether (67 mmol) and 13.8 g p-toluenesulfonic acid monohydrate (73 mmol). The suspension was fitted to with a Dean and Stark water trap and heated to reflux until the calculated amount of water was collected (2.4 mL). The mixture was concentrated to dryness under reduced pressure to yield a yellowish oil. 250 mL DCM was added and the solution washed with 2 × 150 mL saturated aqueous sodium carbonate solution and once with 100 mL water, dried over Na_2SO_4 and concentrated to yield a yellowish oil. Yield: 15.6 g (75%).



$[\alpha]_D^{20}$ = +28.1 (c = 0.006, MeOH). MS (ESI): m/z: 312.2 $[M + H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M + H]^+$: 312.18.

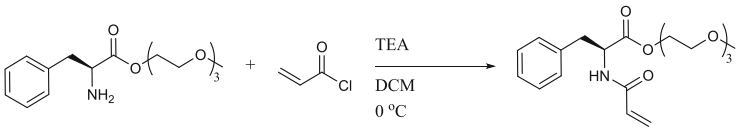
^{1}H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 2.19 (2H(1), m), 2.96 (1H(3b), dd, J = 13.5,7.6 Hz), 3.15 (1H(3a), dd, J = 13.5,5.3 Hz), 3.40 (3H(15), s), 3.57 (2H (14), m), 3.68 (6H(11-13), m), 3.71 (2H(10), m), 3.83 (1H(2), broad-t, m), 4.30 (2H (9), t, J = 4.8 Hz), 7.25 (2H(5), m), 7.26 (1H(7), m), 7.32 (2H(6), m)

^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 40.7 (C-3), 55.6 (C-2), 59.0 (C-15), 64.0 (C-9), 69.1 (C-10), 70.7 (C-11-13), 72.0 (C-14), 127.0 (C-7), 128.8 (C-6), 129.4 (C-5), 137.1 (C-4), 174.7 (C-8).

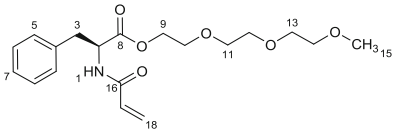
Preparation of (triethylene glycol methyl ether)-D-phenylalaninate

As for (triethylene glycol methyl ether)-L-phenylalaninate described above.
$[\alpha]_D^{20} = -17.1$ (c = 0.007, MeOH).

Preparation of (triethylene glycol methyl ether)-acryloyl-L-phenylalaninate



4.5 g (Triethylene glycol methyl ether)-L-phenylalaninate (15 mmol) and 2.4 mL TEA (18 mmol) were added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 1.3 mL acryloyl chloride (16 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2 × 100 mL 0.5 M HCl and once with 50 mL water, dried over Na_2SO_4 and concentrated to a yellowish oil. Freeze dried to yield a white solid. Yield: 4.5 g (84%).



$[\alpha]_D^{20} = +33.2$ (c = 0.004, MeOH). MS (ESI): m/z: 366.2 $[M + H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M + H]^+$: 366.19.

^{1}H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 3.19 (2H(3), m), 3.35 (3H(15), s), 3.53 (2H(14), m), 3.64 (6H(11-13), m), 3.68 (2H(10), m), 4.28 (2H(9), m), 4.99 (1H(2), dt, J = 7.7,5.6 Hz), 5.65 (1H(18b), dd, *J* = 10.3,1.3 Hz), 6.09 (1H(17), dd, J = 17.0,10.3 Hz), 6.15 (1H(1), d, J = 7.8 Hz), 6.28 (1H(18a), dd, J = 16.9,1.4 Hz), 7.12 (2H(5), dd, J = 7.8,1.2 Hz), 7.23 (1H(7), m), 7.26 (2H(6), m)

^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 37.7 (C-3), 53.1 (C-2), 58.9 (C-15), 64.6 (C-9), 68.8 (C-10), 70.6 (C-11-13), 71.9 (C-14), 127.1 (C-7), 127.2 (C-18), 128.7 (C-6), 129.3 (C-5), 130.5 (C-17), 135.7 (C-4), 164.8 (C-16), 171.5 (C-8).

Preparation of (triethylene glycol methyl ether)-acryloyl-D-phenylalaninate

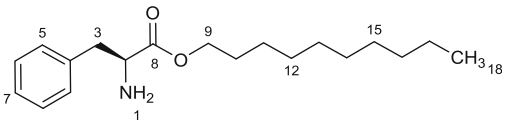
As for (triethylene glycol methyl ether)-acryloyl-L-phenylalaninate described above.
$[\alpha]_D^{20} = -24.7$ (c = 0.003, MeOH).

Preparation of poly-(triethylene glycol methyl ether)-acryloyl-L-phenylalaninate/*N*,*N*-dimethyl acrylamide (p-TEGPhe/DMAA)

0.533 g of (Triethylene glycol methyl ether)-acryloyl-L-phenylalaninate (1.5 mmol) was dissolved in 0.7 mL $CDCl_3$. 0.2 mL *N,N*-dimethylacrylamide (0.1 mmol) and 3.4 mg *N,N*-methylenbisacrylamide (0.01 mmol) was added. 1.5 mg AIBN (0.001 mmol) was added and N_2 bubbled through the solution for 5 min to remove O_2. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and N_2 backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14–16 h) to polymerize.

Preparation of decyl-L-phenylalaninate

10 g L-phenylalanine (61 mmol) was suspended in 250 mL toluene, followed by 12.7 mL decanol (67 mmol) and 13.8 g p-toluenesulfonic acid monohydrate (73 mmol). The suspension was fitted to with a Dean and Stark water trap and heated to reflux until the calculated amount of water was collected (2.3 mL). The mixture was concentrated to dryness under reduced pressure to yield a yellowish oil. 250 mL DCM was added and the solution washed with 2 × 150 mL saturated aqueous sodium carbonate solution and once with 100 mL water, dried over Na_2SO_4 and concentrated to yellowish oil. Re-crystalized from EtOH twice. Yield: 10.0 g (HCl salt) (48.3%).



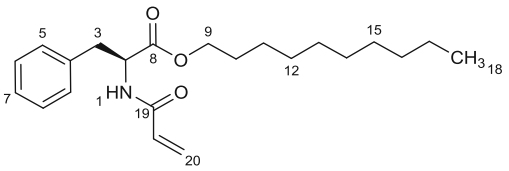
$[\alpha]_D^{20}$ = +14.7 (c = 0.008, $CHCl_3$). MS (ESI): m/z: 306.2 $[M + H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M + H]^+$: 306.24.

^{1}H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 0.88 (3H(18), t, *J* = 7.0), 1.19-1.35 (14H(11-17), m), 1.59 (2H(10), m), 2.53 (2H(1), broad s), 2.94 (1H(3a), dd, *J* = 13.7,7.7), 3.11 (1H(3b), dd, *J* = 13.6,5.7), 3.80 (1H(2), dd, *J* = 7.3,5.6), 4.09 (2H(9), t, J = 6.7 Hz), 7.17-7.33 (5H(5-7), m)

^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 14.2 (C-18), 22.4 (C-17), 25.7, 29.3, 31.8 (C-11-6), 28.4 (C-10), 40.3 (C-3), 55.4 (C-2), 65.2 (C-9), 126.8, 128.5, 129.2 (C-5-7), 136.6 (C-4), 174.1 (C-8)

Preparation of decyl-acryloyl-L-phenylalaninate

1.5 g decyl-L-phenylalaninate (4.9 mmol) and 0.82 mL TEA (5.9 mmol) were added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 0.44 mL acryloyl chloride (5.4 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2 × 100 mL 0.5 M HCl and once with 50 mL water, dried over Na_2SO_4 and concentrated to white solid. Yield: 1.77 g (95.2%).



$[\alpha]_D^{20}$ = +17.2 (c = 0.005, $CHCl_3$). MS (ESI): m/z: 360.3 $[M + H]^+$. m/z calcd for $C_{22}H_{34}NO_3$ $[M + H]^+$: 360.25.

^{1}H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 0.89 (3H(18), t, *J* = 7.1), 1.20-1.35 (14H(11-17), m), 1.59 (2H(10), m), 3.17 (2H(3), m), 4.11 (2H(9), m), 4.96 (1H(2), dt, J = 7.7,5.8 Hz), 5.68 (1H(20a), dd, J = 10.3,1.4 Hz), 6.02 (1H(1), d, J = 7.4 Hz), 6.09 (1H(19), dd, J = 17.0,10.3 Hz), 6.29 (1H(20b), dd, J = 16.9,1.3 Hz), 7.09 (2H(5), m), 7.22-7.31 (3H(6-7), m)

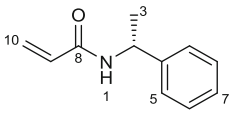
^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 13.8 (C-18), 22.4 (C-17), 25.7, 29.3, 31.8 (C-11-6), 28.3 (C-10), 37.9 (C-3), 52.9 (C-2), 65.8 (C-9), 127.1 (C-20), 127.2, 128.5 (C-6-7), 129.2 (C-5), 130.3 (C-19), 135.7 (C-4), 164.6 (C-16), 171.8 (C-8).

Preparation of poly-decyl-acryloyl-L-phenylalaninate/*N*,*N*-dimethyl acrylamide (p-C_{10}Phe/DMAA)

1.93 g of decyl-acryloyl-L-phenylalaninate (5.4 mmol) was dissolved in 1.5 mL $CDCl_3$. 1.5 mL *N,N*-dimethylacrylamide (14.6 mmol) and 13.0 mg *N,N*-methylenbisacrylamide (0.08 mmol) was added. 2.9 mg AIBN (0.002 mmol) was added and N_2 bubbled through the solution for 5 min to remove O_2. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and N_2 backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14–16 h) to polymerize.

Preparation of (*R*)-*N*-(1-phenylethyl)-acrylamide

2.86 g (*R*)-1-phenylethylamine (23.5 mmol) and 3.9 mL TEA (28.3 mmol) was added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 2.11 mL acryloyl chloride (25.9 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2 × 100 mL 0.5 M HCl and once with 50 mL water, dried over Na_2SO_4 and concentrated *in vacuo* to yield a white solid. Yield: 4.03 g (97.6%).



$[\alpha]_D^{20}$ = +22.2 (c = 0.009, MeOH). MS (ESI): m/z: 176.1 $[M + H]^+$. m/z calcd for $C_{11}H_{14}NO$ $[M + H]^+$: 176.11.

^{1}H NMR (DMSO-d_6, 400 MHz, 298 K): δ = 1.38 (3H(3), d, *J* = 7.3), 5.0 (1H (2), p, *J* = 7.5), 5.60 (1H(10a), dd, *J* = 10.1,2.2), 6.08 (1H(10b), dd, *J* = 17.1,2.2),

6.29 (1H(9), dd, *J* = 17.1,10.2), 7.21 (1H, m), 7.29-7.39 (4H(5–6), m), 7.76 (1H(1), d, J = 7.8 Hz)

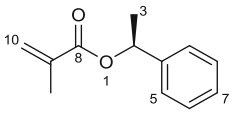
^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 22.7 (C-3), 48.2 (C-2), 125.7 (C-10), 126.4 (C-5), 127.2 (C-7), 128.7 (C-6), 132.2 (C-9), 144.9 (C-4), 163.9 (C-8)

Preparation of poly-(*R*)-*N*-(1-phenylethyl)-acrylamide/*N*,*N*-dimethyl acrylamide (p-PhEtN/DMAA)

0.23 g of (*R*)-*N*-(1-phenylethyl)-acrylamide (1.3 mmol) was dissolved in 0.2 mL acetone-d_6. 0.25 mL *N*,*N*-dimethylacrylamide (2.4 mmol) and 2.7 mg *N*,*N*-methylenbisacrylamide (0.02 mmol) was added. 2.8 mg AIBN (0.002 mmol) was added and N_2 bubbled through the solution for 5 min to remove O_2. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and N_2 backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14–16 h) to polymerize.

Preparation of (*S*)-1-phenylethyl methacrylate

2.0 mL (*S*)-1-phenylethanol (16.5 mmol) and 2.8 mL TEA (19.9 mmol) were added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 1.8 mL methacryloyl chloride (18.2 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2 × 100 mL 0.5 M HCl and once with 50 mL water, dried over Na_2SO_4 and concentrated *in vacuo* to yield a reddish oil. Yield: 2.54 g (80.5%).



$[\alpha]_D^{20}$ = −29.4 (c = 0.008, MeOH). MS (ESI): m/z: 191.1 $[M + H]^+$ m/z calcd for $C_{12}H_{15}O_2$ $[M + H]^+$: 191.11.

^{1}H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 1.50 (3H(3), d, *J* = 6.7), 1.89 (3H (11), broad s), 5.49 (1H(10a), p, *J* = 1.7), 5.87 (1H(2), q, *J* = 6.6), 6.08 (1H(10b), dq, *J* = 1.6,0.8), 6.29 (1H(9), dd, *J* = 17.1,10.2), 7.21 (1H, m), 7.24-7.32 (4H(5–6), m)

^{13}C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 18.3 (C-11), 22.3 (C-3), 72.4 (C-2), 125.4 (C-10), 126.0 (C-5), 127.8 (C-7), 128.5 (C-6), 136.6 (C-9), 141.7 (C-4), 166.6 (C-8)

Preparation of poly-(*S*)-*O*-(1-phenylethyl)-methacrylate (p-PhEtO)

0.57 g of (*S*)-1-phenylethyl methacrylate (3.0 mmol) was dissolved in 0.17 mL acetone-d_6. 1.4 µL EGDMA (0.007 mmol) was added. 1.3 mg AIBN (0.001 mmol) was added and N_2 bubbled through the solution for 5 min to remove O_2. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and N_2 backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14–16 h) to polymerize.

Swelling of polymers

See Sect. 6.4 regarding polymers from 3 mm NMR tubes. If the inserted plug visibly moved during swelling the sample was excluded from the study of chiral polymers, due to an observed large error between alignments of equal enantiomers when this happened. RDCs could still be extracted though so this is not a problem for achiral polymers.

Also see Sect. 6.4 for information regarding synthesis of the PDMAA polymer. All other necessary information is available in figures and tables throughout this chapter.

NMR experiments

The CLIP-HSQC experiment was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. The concentration of analytes was ~150 mM for isotropic samples and ~220 mM (strychnine) or ~320 mM (IPC/menthol) for aligned samples.

Simulations

Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations [25], using the program MacroModel version 10.8 [26, 27]. The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization [26, 28].

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations [29]. Structures were optimized to a B3LYP/6-31(d) level of theory.

References

1. I.W. Wainer, Hum. Psychopharmacol. Clin. Exp. **16**, S73–S77 (2001)
2. N.M. Maier, P. Franco, W. Lindner, J. Chromatogr. A **906**, 3–33 (2001)
3. O. Mcconnell, A.B. Ii, C. Balibar, N. Byrne, Y. Cai, G.U.Y. Carter, M. Chlenov, L.I. Di, K. Fan, I. Goljer, Y. He, D.O.N. Herold, M. Kagan, E. Kerns, F. Koehn, C. Kraml, V. Marathias, B. Marquez, L. Mcdonald, L. Nogle, C. Petucci, G. Schlingmann, G. Tawa, M. Tischler, R.T. Williamson, A. Sutherland, W. Watts, M. Young, M. Zhang, Y. Zhang, D. Zhou, D. Ho, Chirality **19**, 658–682 (2007)
4. J. Clayden, N. Greeves, S. Warren, P. Wothers, Organic Chemistry (OUP Oxford, 2001)
5. T.J. Wenzel, C.D. Chisholm, Prog. Nucl. Magn. Reson. Spectrosc. **59**, 1–63 (2011)
6. T.J. Wenzel, C.D. Chisholm, Chirality **23**, 190–214 (2011)
7. R.C. Breton, W.F. Reynolds, Nat. Prod. Rep. **30**, 501–524 (2013)
8. D. Parker, Chem. Rev. **91**, 1441–1457 (1991)
9. F. Balzano, T. Brotin, J.H. redehöft, J.P. Dutasta, A.C. Evans, C. Giri, L. Guy, S.V. Hoffmann, N.C. Jones, K. Manoli, M. Magliulo, A. Martinez, C. Meinert, U.J. Meierhenrich, A. E. Sorochinsky, V.A. Soloshonok, L. Torsi, O. Trapp, G. Uccello-Barretta, T.J. Wenzel, Differentiation of Enantiomers II, 1st edn. ed. by V. Schurig (Springer: Heidelberg, 2013)

10. R. Bhushan, H. Brückner, Amino Acids **27**, 231–247 (2004)
11. P. Marfey, Carlsberg Res. Commun. **49**, 591–596 (1984)
12. A. Krupp, M. Reggelin, Magn. Reson. Chem. **50**, S45–S52 (2012)
13. M. Schmidt, H. Sun, A. Leonov, C. Griesinger, U.M. Reinscheid, Magn. Reson. Chem. **50**, S38–S44 (2012)
14. H. Yuki, K. Ohta, K. Uno, S. Murahashi, J. Polym. Sci. A **6**, 829–841 (1968)
15. T. Kitayama, W. Shibuya, K. Katsukawa, Polym. J. **34**, 405–409 (2002)
16. N.A. Porter, T.R. Allen, R.A. Breyer, J. Am. Chem. Soc. **114**, 7676–7683 (1992)
17. N.A. Porter, I.J. Rosenstein, R.A. Breyer, J.D. Bruhnke, W.X. Wu, A.T. McPhail, J. Am. Chem. Soc. **114**, 7664–7676 (1992)
18. K. Hatada, T. Kitayama, K. Ute, Prog. Polym. Sci. **13**, 189–276 (1988)
19. N.C. Meyer, A. Krupp, V. Schmidts, C.M. Thiele, M. Reggelin, Angew. Chemie Int. Ed. **51**, 8334–8338 (2012)
20. C. Gayathri, N.V. Tsarevsky, R.R. Gil, Chem. Eur. J. **16**, 3622–3626 (2010)
21. T. Montag, C.M. Thiele, Chem. Eur. J. **19**, 2271–2274 (2013)
22. R. Berger, J. Courtieu, R.R. Gil, C. Griesinger, M. Köck, P. Lesot, B. Luy, D. Merlet, A. Navarro-Vázquez, M. Reggelin, U.M. Reinscheid, C.M. Thiele, M. Zweckstetter, Angew. Chemie Int. Ed. **51**, 8388–8391 (2012)
23. A.O. Frank, J.C. Freudenberger, A.K. Shaytan, H. Kessler, B. Luy, Magn. Reson. Chem. **53**, 213–217 (2015)
24. W. Zong, G. Li, J. Cao, X. Lei, M. Hu, H. Sun, C. Griesinger, R.X. Tan, Angew. Chemie Int. Ed. **55**, 3690–3693 (2016)
25. Maestro (Schrödinger, LLC, New York, 2015)
26. Schrödinger (LLC, New York, NY 2016)
27. Desmond (D. E. Shaw Research, New York, 2015)
28. G. Chang, W.C. Guida, W.C. Still, J. Am. Chem. Soc. **111**, 4379–4386 (1989)
29. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in Gaussian 09, Revision B.01 (Gaussian Inc., Wallingford, CT, 2010)

Chapter 8
Tensor Free RDC Calculations

For the theoretical basis of RDCs, the reader is referred to Chap. 5. The theory behind the actual back-calculations of RDCs from experimental data and 3D structures will be described briefly below to introduce a new method of back-calculation. The introduction will focus entirely on singular value decomposition (SVD), since this is by far the most utilized method in small molecule RDC calculations. To support a possible need for a new back-calculation method, it is important to understand the basics of SVD and the advantages and disadvantages involved when using this method.

8.1 Singular Value Decomposition

Losonczi et al. described the mathematical basis of the SVD method for back-calculation of RDCs of macromolecules [1]. Here the principles of SVD are included in Eqs. (8.1)–(8.12).

Tensors in RDC calculations are based on an expanded form of Eq. (5.2), Sect. 5.1. Here, the alignment frame is allowed to shift in comparison to the direction of the magnetic field (usually chosen as the z-axis) and now includes an axial (A_a) and rhombic (A_r) component, see Eq. (8.1) [2–5]

$$D_{IS} = -\frac{\hbar\gamma_I\gamma_S\mu_0}{16\pi^2 r_{IS}^3}\left[A_a\left(3\cos^2\theta_{IS} - 1\right) + \frac{3}{2}A_r\sin^2\theta_{IS}\cos 2\varphi_{IS}\right] \tag{8.1}$$

where γ is the gyromagnetic ratio, $\hbar$ is the Planck constants divided by 2π, μ_0 is the vacuum permeability constant, r is the cubed distance between the nuclei and θ is the angle between the internuclear vector and the magnetic field. φ is the azimuthal angle that describes the projection of the RDC vector to the xy-plane with respect to the x-axis, as illustrated Fig. 8.2, left.

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_8

This will lead to an averaging of the internuclear vectors positioned along the x-axis and is needed to unambiguously place the molecules in the alignment frame. Four possible alignments are always present due to equal theoretical RDCs of the alignments as illustrated in Fig. 8.1.

In Fig. 8.2 left, the direction of the magnetic field is along the z-axis, and the length of the *a* vector will serve as the degree of alignment (A_a). But the direction of the magnetic field will also be along the x-axis to some degree, and the length of *b* determines how much it is aligned here (A_r). The probability, or degree, of the shift is given by the rhombicity (R) given in Eq. (8.1).

Without introducing the rhombicity, the theoretical fit between experimental and back-calculated data would be worse for most alignments since some averaging is present in the experimental data, and the compound would not be specifically placed in space but only by an angle to the magnetic field with free rotation around the magnetic field vector. An experimental dataset which back-calculates to a structure with a low rhombicity is thus averaged less along the x-axis than a structure with a high rhombicity.

$$A_a = \frac{S_{zz}}{2} \tag{8.2}$$

$$A_r = \frac{1}{3}\left(S_{xx} - S_{yy}\right) \tag{8.3}$$

$$R = \frac{A_r}{A_a} = \frac{2}{3}\frac{S_{xx} - S_{yy}}{S_{zz}} \tag{8.4}$$

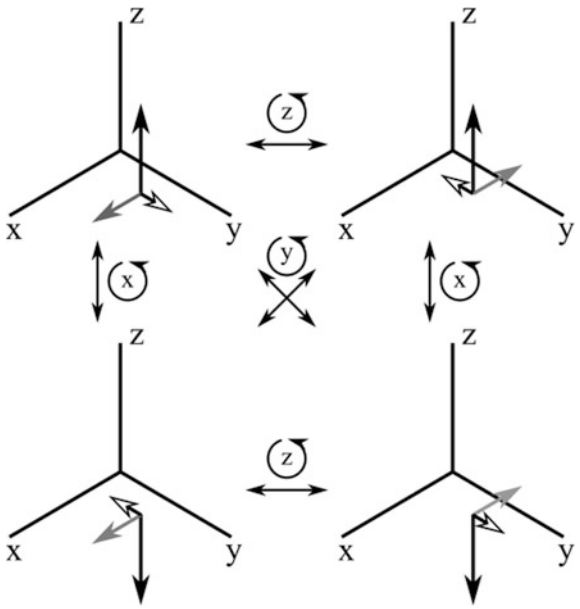


Fig. 8.1 The four positions of a molecule (here represented by three vectors) which will give equal RDC values from tensor RDC calculations. All of the indicated rotations are 180° around the axis

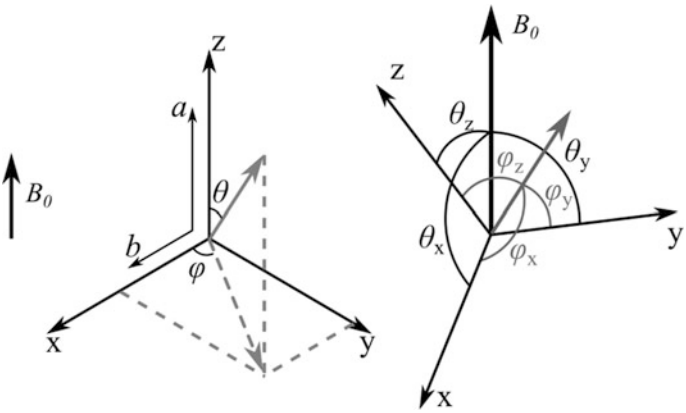


Fig. 8.2 Left: The scaling vectors and angles involved in the extended formula for tensor RDC calculations, Eq. (8.1). Right: The molecular axis compared to the direction of the magnetic field, including the angles used in SVD fitting. The angles θ are the angles between the molecular axis and the magnetic field, and the angles φ are the angles between an internuclear vector and the molecular axis, see Eqs. (8.1) to (8.12)

To establish a tensor the Eqs. (8.1)–(8.4) are used to produce an order matrix as a representation of the anisotropic averaging leading to dipolar couplings, given in Eq. (8.5) and (8.6) Fig. 8.2, right is a good reference to keep track of the angles used in the following.

$$A = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \tag{8.5}$$

$$S_{ij} = \left\langle 3\cos\theta_i \cos\theta_j - k_{ij} \right\rangle \tag{8.6}$$

where k_{ij} is the Kronecker delta ($k_{ij} = 1$ if $i = j$ and $k_{ij} = 0$ if $i \neq j$). The matrix is symmetric ($S_{ij} = S_{ji}$) and traceless ($S_{xx} + S_{yy} + S_{zz} = 0$). Thus only five elements are needed to obtain the full matrix A, chosen to be S_{yy}, S_{zz}, S_{xy}, S_{xz} and S_{yz}. The five elements translate to three angles (which need to be non-parallel) the alignment strength and the rhombicity. The Eqs. (8.5) and (8.6), if used directly on a 3D structure, suppose a situation where the alignment frame and the molecular frame have an identical orientation which is not generally true. To correlate the molecular frame to the alignment frame, the equation is thus extended into Eq. (8.7) and further to (8.8) where φ_i is the angle between internuclear vector and the molecular frame.

$$D^m = D^m_{max} \sum_{ij=x,y,z} S_{ij} \cos\varphi_i^m \cos\varphi_j^m \tag{8.7}$$

$$D_{red}^{m} = \frac{D^{m}}{D_{max}^{m}} = \sum_{ij=x,y,z} S_{ij} \cos\varphi_i^m \cos\varphi_j^m \tag{8.8}$$

where D_{max}^{m} is $-\mu_0 \gamma_I \gamma_S \hbar r^{-3}/(16\pi^2)$. This equation is then used to setup a set of linear equations, as seen in (8.9). Note that the φ angles are always known from a 3D structure and the matrix is easily established.

$$\begin{bmatrix} D_{red}^{1} \\ \vdots \\ D_{red}^{n} \end{bmatrix} = \begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix}$$

$$\begin{bmatrix} \cos^2\varphi_y^1 - \cos^2\varphi_x^1 & \cos^2\varphi_z^1 - \cos^2\varphi_x^1 & 2\cos\varphi_x^1\cos\varphi_y^1 & 2\cos\varphi_x^1\cos\varphi_z^1 & 2\cos\varphi_y^1\cos\varphi_z^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos^2\varphi_y^n - \cos^2\varphi_x^n & \cos^2\varphi_z^n - \cos^2\varphi_x^n & 2\cos\varphi_x^1\cos\varphi_y^n & 2\cos\varphi_x^1\cos\varphi_z^n & 2\cos\varphi_y^1\cos\varphi_z^n \end{bmatrix} \tag{8.9}$$

For $n > 5$ the system will be overdetermined, while it will be unsolvable for $n < 5$. As a consequence, RDCs of five non-parallel vectors are needed to describe a system using this method. If $n = 5$ the system is solvable as five equations with five unknowns, but this situation is not particularly interesting as one could in principle fit almost anything to "perfection". Thus SVD was introduced as a mean to solve the overdetermined set of linear equations, and to find the order tensor with the best overall fit to the data in a least square sense.

It is used that the $N \times M$ matrix $[B]$ in Eq. (8.10), from the matrix in (8.9), may be described by the product of an $M \times N$ matrix $[U]$, an $N \times N$ diagonal matrix $[W]$, with non-negative diagonal elements, and the transpose of a $N \times N$ orthogonal matrix $[V]$, which allows the definition of $[B]^{-1}$ as can be seen in Eq. (8.11).

$$\begin{bmatrix} D_{red}^{1} \\ \vdots \\ D_{red}^{n} \end{bmatrix} = \begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} [B] \tag{8.10}$$

$$[B] = [U] \begin{bmatrix} \omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \omega_n \end{bmatrix} \left[V^T\right] \leftrightarrow [B]^{-1} = [V][diag(1/\omega_i)][U]^T \tag{8.11}$$

This may then be used to solve Eq. (8.9) as Eq. (8.12).

$$\begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} = \begin{bmatrix} D^1_{red} \\ \vdots \\ D^n_{red} \end{bmatrix} [V][diag(1/\omega_i)][U]^T \tag{8.12}$$

SVD is implemented as a function in mathematical programs such as Matlab®, making the decomposition easy [6].

The result is a least square fit of the experimental data to the 3D structure, which will rotate the structure into the alignment frame and scale the RDC values according to the experimental data [1]. It is thus a back-calculation method where a 3D structure is needed to determine the order tensors. For rigid structures SVD works very well and is a powerful, albeit simple, tool in macromolecular and small molecule work alike [5, 7–9].

The drawback of this method is evident only if investigating flexible compounds. The arising problems are caused by establishing order tensors in the first place and are not a flaw in the SVD fitting as such. Since order tensors will determine a rotation of the structure the implementation for multiple structures is not straightforward but may still be approached. In general two methods have been applied; the multi conformer multi tensor (MCMT) and multi conformer single tensor (MCST) methods [8]. The MCMT method builds on the assumption that one may describe the orientation of flexible systems by obtaining order tensors for each conformer. As this method involves multiple order tensors, $5n$ RDCs are needed *per conformation*, with n being the number of conformers of the system [1]. This approach is thus rarely feasible for flexible small molecules [1, 8].

8.1.1 Multi Conformer Single Tensor

The MCST method assumes that a single tensor is sufficient to describe the alignment of all conformers [8]. In other words it is assumed, that all conformers are aligned equally in space—or at least parts of the conformers. This may be a good approximation in certain situations, when the structural flexibility is limited to small groups in regard to an overall structure or to small movements. This is an obvious limitation, since a reference frame needs to be established and all structures are linked to this reference frame when rotated into the alignment frame [8]. The resulting back-calculated RDCs are thus dependent on the original overlay of the input structures, and parts of the molecule need to be identified that are independent in the orientation in space between the conformers. While a rigid part may be identified for the majority of organic compounds (if not a different set of problems will be more pressing in 3D structure determination) it may be harder to identify a part of a molecule that, independent of conformation, is situated evenly in space compared to the rest of the structure. A general approach to identify the optimal overlay of structures has yet to be reported. An approach could be to overlay

vectors with the largest absolute RDCs, as these are probably the most static in their orientation to B_0. Whether this is better than e.g. identifying large static surfaces of the compound is not known, and was not thoroughly investigated.

The methodology is thus limited to specific cases and over-interpretation is deceptively easy, which will be showed *vide infra*. It may be implemented in different ways; here the focus will on be the implementation utilized in the program MSPIN [2]. Here the population is iteratively fitted to the experimental RDCs, by constructing alignment tensors from the populations and the evaluating the correlation of experimental and calculated RDCs.

8.1.2 RDCs of Methylene and Methyl Groups

When dealing with overlying resonances such as methylene groups that do not split and methyl groups, various methods have been proposed to back-calculate the RDCs.

$^1D_{\mathrm{CH3}}$ values are usually obtained from spectra and, while not immediately useful, are converted to the associated $^1D_{\mathrm{CCH3}}$ coupling constant, which is easier to implement in the given types of back-calculations. This is achieved by assuming a 3-jump model and that the methyl group is a perfect tetrahedral in Eq. (8.13) [10, 11]. This assumption makes a conversion from a theoretical CH vector to the actual CC vector possible using Eq. (8.14).

$$^1D_{CH_3} = \frac{3\cos^2\varphi - 1}{2} D_{\parallel} = -\frac{1}{3} D_{\parallel} \tag{8.13}$$

$$^1D_{CCH_3} = \frac{\gamma_C}{\gamma_H} \frac{r_{CH}^3}{r_{CC}^3} D_{\parallel} \rightarrow D_{\parallel} = -\frac{3\gamma_C}{\gamma_H} \frac{r_{CH}^3}{r_{CC}^3} {}^1D_{CH_3} \tag{8.14}$$

where φ is the angle of the CH vector and the rotation axis of the methyl group, $D_{\parallel}$ is a virtual CH vector pointing in the direction of the CC vector, γ is the gyromagnetic ratio and r is the length of the corresponding bonds.

For methylene groups it has been shown that if an average coupling constant is extracted, the data may still be used in the fitting of experimental data to 3D structures, by calculating the theoretical average from the individual CH vectors [12].

8.2 Tensor Free Calculations of RDCs

Tensor free calculation of RDCs, also dubbed the 𝔍-method, has been shown to work well for macromolecules but has not directly been adopted in the small molecule community [13, 14]. In a publication by Luy et al. in 2015, a resembling although different methodology is utilized [15]. For macromolecules the 𝔍-method

is used implicitly in the 3D structure generation as constraints [13, 14]. This is not necessarily wanted for small molecules due to similar reasons as discussed in e.g. Section 3.3.1 for NOE calculations: In short, the use of constrains may limit or disturb the conformational space.

We thus set out to translate the method into one more suitable for conformational investigation of small molecules. In the following description this method is called the θ-method, for distinction and recognition of inspiration [13, 14].

8.3 The θ-Method

The θ-method is simplistic in its origin; it utilizes only the standard equation for dipolar couplings, Eq. (8.15) reiterated from Eq. (5.2) Sect. 5.1, with no tensors to correlate the alignment frame to the molecular frame, as shown in Fig. 8.2.

$$D_{IS} = -\frac{\hbar\gamma_I\gamma_S\mu_0}{16\pi^2}\left\langle \frac{1}{r_{IS}^3}\left(3\cos^2\theta_{IS} - 1\right)\right\rangle \tag{8.15}$$

This equation may be shortened to (8.16).

$$D_{IS} = D_{max}^{IS} D_a \left\langle 3\cos^2\theta_{IS}(R) - 1\right\rangle = D_{max}^{IS} D_a \int P(R)\left(3\cos^2\theta_{IS}(R) - 1\right)dR \tag{8.16}$$

R covers external (rotational) and intramolecular (conformational) motion, D_a is the degree of alignment and κ is the constants of (8.16) assuming that only $^1D_{\mathrm{CH}}$ data is used and r is constant. If this is not the case the equation is expanded to (8.17).

$$\begin{aligned} D_{IS} &= D_{r,max}^{IS} D_a \left\langle \frac{3\cos^2\theta_{IS}(R) - 1}{r(R)^3}\right\rangle \\ &= D_{r,max}^{IS} D_a \int P(R)\left(\frac{3\cos^2\theta_{IS}(R) - 1}{r(R)^3}\right)dR \end{aligned} \tag{8.17}$$

This may in principle be solved if all conformers and a rotational preference are known. Luy et al. showed that this approach worked on a system where strychnine was simulated in the presence of a polymer strand which induced overall alignment [15].

Instead of including explicit orienting media in the simulations or orienting the molecular frame to an alignment frame, a novel approach was taken; all conformers from a simulation are rotated in space, using rotation matrices as in (8.18).

$$Rot_z(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, Rot_x(\omega) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix} \quad (8.18)$$

To get full rotational sampling in 3D space compared to one axis (here z), a combination of only two matrices are needed; x and y, x and z or y and z (for similar reasons as outlined in Fig. 8.1). Here, x and z were chosen, so that the first rotation is around the z axis, which is also the chosen B_0 axis, and the second rotation is around the x axis, as illustrated in Fig. 8.3. In practice, one of the rotations only needs to be 180° as to avoid repeating magnetic field angles, while the other is 360°.

The angle of the RDC vectors from the rotated coordinates to the magnetic field was then determined along with the length of the vector, which in most cases is constant since $^1D_{CH}$-coupling constants are used. From the length and angle RDCs of all conformers and rotations are easily calculated after D_a is determined, *vide infra*. The RDC data sets of the rotated structures are then fitted to the data iteratively, by minimizing the Q-factor (Q) in (8.19), using only the conformations and rotations that lead to an increased correlation of the found average RDCs to the experimental.

$$Q = \frac{\sum_i D_i^{calc} - D_i^{exp}}{\sum_i D_i^{exp}}, D_i^{calc} = \frac{\sum_n D_{i,n}^{calc}}{n} \quad (8.19)$$

Thus, all conformers may be aligned differently, if the data permits. This is at a first glance a possibly problematic approach with multiple open questions:

1. Would the approach give reasonable fits between experimental data and 3D structures?
2. Would the calculation procedure make it possible to discern stereoisomers?
3. And most importantly, will this method fit populations of flexible compounds?

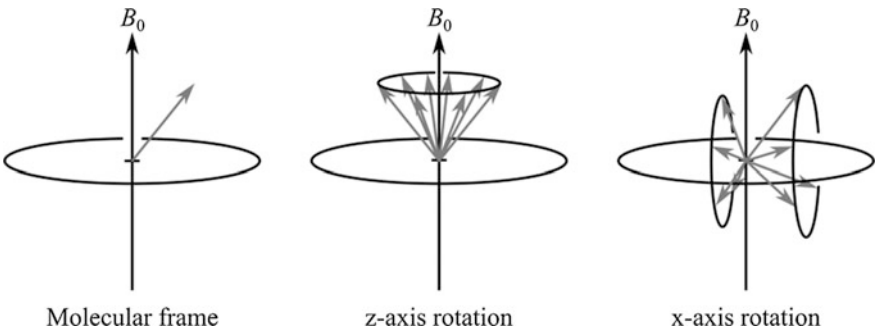


Fig. 8.3 An illustration showing the effect of z- and subsequent x-rotation on an internuclear vector. While the angle to the magnetic field is equal for all vectors after z-rotation, the angle is different for the vectors after equal x-rotation (unless a vector is parallel to z-axis)

The easiest way to verify the method is to test it on a compound where the answer is already known and return to rigid compounds to establish prove of concept. Multiple tests were thus conducted on multiple data sets of strychnine, *vide infra*.

If long-range RDCs were needed the length of the vector of the different conformations needs to be calculated and averaged as well. This was in practice easily implemented by incorporating the distance information into the theoretical RDCs, as in Eq. (8.17).

8.4 Estimating D_a

As mentioned above the strength of alignment needed to be determined or estimated. Here inspiration from macromolecular literature was used [16]. The difference of the approach taken here lies in the fact that D_a is not used for tensor determination and is not split into an axial (D_a) and a rhombic component (D_r) and the rhombicity is thus not estimated as it is not needed due to implicit inclusion in tensor free calculations. Even though the two D_a parameters are used differently, their contribution to the calculations is equal in giving the maximum possible values of the RDCs, and they may be cautiously compared.

Initially it is assumed that the vectors in the molecules are non-parallel and of equal length (neither assumptions are strictly necessary, but simplify the following)[1], and that at least one random vector is either parallel or perpendicular to the magnetic field, B_0. That particular vector would then have the minimum ($D_{exp,min}$) or maximum ($D_{exp,max}$) experimental RDC respectively from Eqs. (8.15) or (8.16). From the equation it is evident that the equations in (8.20) are true, as the constant k covers all other contributions to the RDCs than the angle to the magnetic field.

$$\begin{aligned} D_{exp,min} &= -k\left(3\cos^2\theta_{IS} - 1\right) = -k\left(3\cos^2 0 - 1\right) = -2k \\ D_{exp,max} &= -k\left(3\cos^2\theta_{IS} - 1\right) = -k\left(3\cos^2 \pi - 1\right) = 1k \\ D_{exp,max} &= -D_{exp,min}/2 \\ R_D &= -D_{exp,min}/2D_{exp,max} \end{aligned} \tag{8.20}$$

From these equations D_a may be estimated; if $R_D > 1$ then D_a is $-D_{exp,min}/2$, and if $R_D \leq 1$, D_a is $D_{exp,max}$, divided by D_{max} of the given vector. The determined D_a is thus the minimum D_a that could possibly lead to the experimental data. It is crude, but may be refined later if needed by up- or down-scaling D_a. It is assumed that D_a is equal for all conformers. This approximation is actually a prerequisite for

[1]If all vectors are parallel, D_a may not be estimated, but the angles between internuclear vectors (which are already known) may in principle be solved and the molecule will be fitted with all internuclear vectors either parallel of perpendicular to the magnetic field, dependent on the sign of the RDCs.

most RDC back-calculation theory; it translates to assuming that the interaction with the alignment media is independent of conformation, and thus does not influence the conformational average. Since that requirement is necessary to obtain any meaningful conformational data from RDCs this assumption is hopefully upheld. There is a larger possibility of finding a vector that is perpendicular to the magnetic field, easily realized by the fact that the probability of finding a vector with an angle θ to the magnetic field is proportional to $\sin\theta$ [16, 17]. This is generally not considered in the following, since the value of D_a is scaled, but it could be introduced to the calculation.

8.5 Implementation

The θ-method was implemented in Matlab®. The overall methodology of the script is described above or found in Appendix A7. The script handles RDCs, NOEs and *J*-coupling constants, individually or in combination, given input files and a structure file with one or more structures (.pdb, .mol2 or .sdf supported). Examples of the different inputs are also given in Appendix A7.

8.6 Results

In this section the θ-method is utilized to back-calculate RDCs of several compounds and the results are compared to SVD calculations using the MCST approximation when appropriate. But first a couple of questions needed to be answered in order to establish a proof of concept of the method.

8.6.1 Proof of Concept—Strychnine

Several questions needed to be answered, in order to evaluate the applicability of the θ-method, and are addressed below. Since the method is based on existing methods, the proof of concept part is more a study in the limitations than whether the method is possible, though that is also investigated. To achieve proof of concept the structure of strychnine was utilized. Strychnine is rigid with multiple $^1D_{CH}$ vectors, and is thus generally utilized for many purposes. Two datasets of strychnine were investigated with correlating results. One of the sets is used as an example in the following.

Initial fitting of strychnine

The structure of strychnine (Fig. 8.4) was easily fitted to the RDC data using either SVD or the θ-method, as seen in Table 8.1.

Table 8.1 Experimental versus calculated RDCs for the major conformer of strychnine

Nuc 1	Nuc 2	Exp.	θ-method	SVD
C1	H1	26.6	26.3	26.2
C2	H2	13.8	13.7	13.9
C3	H3	23.8	20.1	20.3
C4	H4	28.3	26.6	26.6
C8	H8	−14.4	−14.1	−14.4
C11	H11a	12.3	13.8	13.7
C11	H11b	−25.8	−28.6	−28.9
C12	H12	−28.4	−29.3	−29.3
C13	H13	−4.3	−4.5	−4.7
C14	H14	−27.9	−29.1	−29.0
C15	H15a	7.0	5.6	5.5
C15	H15b	7.5	7.4	7.3
C16	H16	15.2	14.2	14.2
C18	H18a	1.6	−0.4	−0.5
C18	H18b	14.7	13.6	13.6
C20	H20a	−16.0	−14.5	−14.6
C20	H20b	−4.1	−3.4	−3.6
C22	H22	0.7	3.7	3.6
Q−factor			0.092	0.092

Calculated by SVD in MSPIN and the θ-method. Aligned in PDMAA (Polymer 8.1)

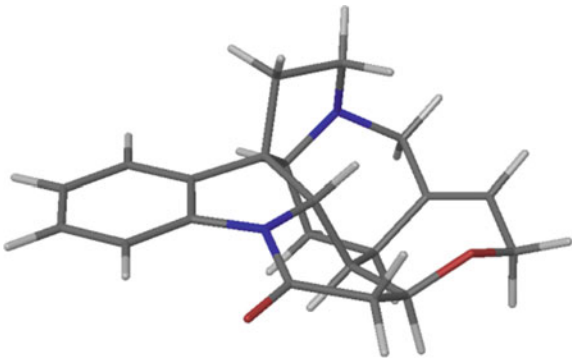


Fig. 8.4 The major conformer of strychnine

The resulting back-calculated RDCs were very much alike with an RMSD of 0.13 Hz and a MAE of 0.06 Hz between the two methods. While this was reassuring the work relied heavily on the following. The general alignment is compared in Fig. 8.5, but as the averaging is implicit in SVD and explicit in the θ-method, the comparison is of only the most abundant alignment.

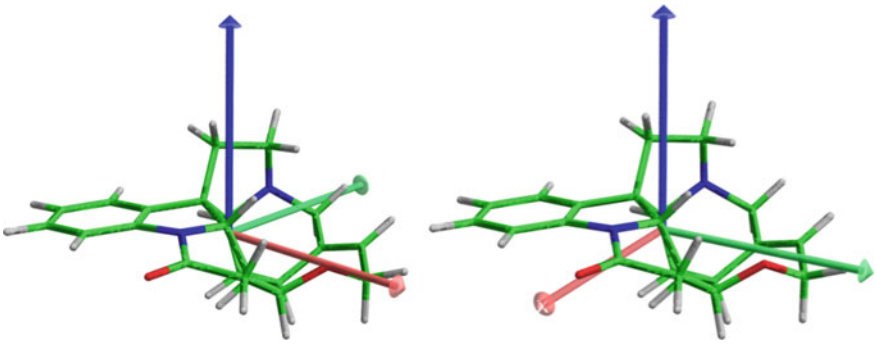


Fig. 8.5 The alignment frame from an SVD fitting of strychnine (left) and the molecular frame of the most abundant alignment to the magnetic field from a θ-method fitting (right). Both alignments have the z-axis aligned with the magnetic field. Note that the x- and y-axis are without influence in the θ-method, and the alignments are almost identical. Aligned in PDMAA (Polymer 8.1)

Rotation angle φ

The first potential drawback of the method is caused by the fact that a rotation must be defined, and thus the back-calculation of RDCs is based on a finite number of rotations. It is not possible to make the rotation angle infinitely small due to computer memory and time of calculations. In comparison the SVD method rotates the structure into the linear best fit with a theoretical infinite rotational resolution. Conditions were tested to establish that it was possible to find a rotation angle which gives reliable results and is computationally viable.

To investigate this, the 3D structure of the major conformer of strychnine was used. When optimized this structure has a random orientation in space and randomly 100 structures were generated from 100 random rotation angles of the structure along the z- and x-axis (using a rotational resolution below 0.0001°). The 100 randomly chosen structures were rotated by 90, 60, 30, 10, 1, 0.5 or 0.1° and back-calculated RDCs compared to the experimental data. The summary of this may be found in Fig. 8.6 and Table 8.2.

The lowest rotation angle used in the further investigation was 0.1° which is a practical decision; to decrease an order of magnitude from 0.1° one would need ~227 GB of RAM, way above standard computer setups (and even clusters).[2] This could probably be alleviated by saving and loading the structures continuously in the Matlab® script, but this was not a feasible solution due to increased calculation times.

It should be addressed that a rotation of 10 or even 30° fitted the data just as well as lower rotational steps. It is unlikely that a rotation of 30° will produce a structure that is situated rightly in the alignment frame, and the generally low *Q-factor* is

[2]This is a result of the script where all rotations of all vectors are saved in matrices. Estimated as RAM(0.1° rot.) × 10 × 10.

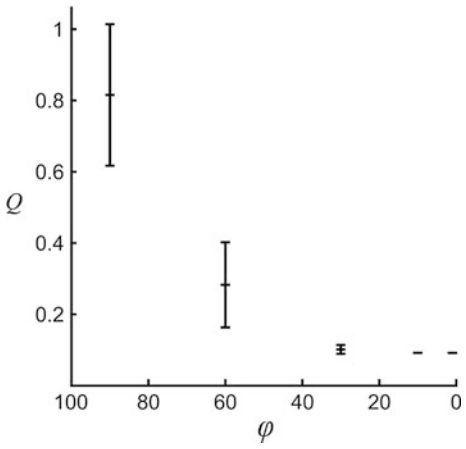


Fig. 8.6 Q-factor of an RDC fitting of strychnine using different rotation steps (φ) in degrees. Mean and standard errors (error bars) shown. Data set used for strychnine is Polymer 8.1

Table 8.2 Statistics of the Q-factor of an RDC fitting of strychnine using different rotation steps (φ) starting from 100 different rotations

φ/°	Mean	St. dev.	Median	Min	Max
–	1.6290	0.4154	1.6950	0.4767	2.2857
90	0.8156	0.1986	0.8910	0.2060	0.9981
60	0.2829	0.1195	0.2835	0.0934	0.5357
30	0.1016	0.0123	0.0963	0.0919	0.1468
10	0.0919	0.0002	0.0919	0.0916	0.0924
1	0.0918	0.0002	0.0917	0.0915	0.0921
0.5	0.0917	0.0002	0.0916	0.0915	0.0921
0.1	0.0917	0.0001	0.0916	0.0915	0.0921

Data set used for strychnine is Polymer 8.1

attributed to the fact that the rotational average is optimized. It is thus possible to construct data from multiple aligned structures which averages in line with the observed data. If a rotation by 1° or lower is utilized, the possibility to be in the alignment frame is vastly increased, and the minor averaging needed may even correlate well to what is actually observed in experiments. Due to the vast increase in computational requirements when decreasing the rotation step size by an order of magnitude, a rotation of 1° was used to perform the back-calculations and fittings in the remainder of this thesis.

D_a estimations

To do the calculations as outlined in (8.16) and (8.17) it was a prerequisite that D_a may be estimated. The theoretical approach was already explained in (8.20). In order to test whether the approach was viable, D_a was scaled by a constant S_{Da} in the fitting to experimental data and the resulting Q-factor was evaluated. An example of this, using strychnine again, is seen in Fig. 8.7.

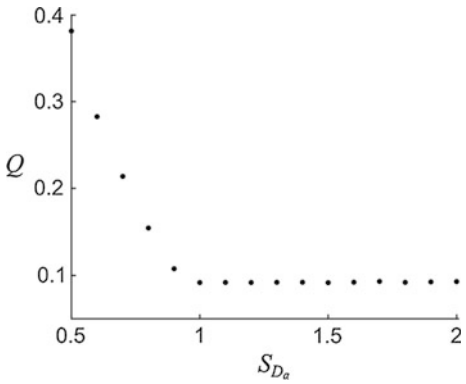


Fig. 8.7 Q-factor of an RDC fitting of strychnine using a D_a obtained by the approach outlined in Sect. 8.4 and scaled by S_{Da}. Data set used for strychnine is Polymer 8.1

The assumptions used in estimating D_a seemed to be valid for strychnine in the experimentally obtained alignment frame, as an increase in D_a did not lead to a better correlation to the experimental data. A decrease in D_a led to a worse correlation, in good correlation to the fact that the larger observed RDCs, independent of sign, are no longer obtainable. The fact that $S_{Da} = 1$ was the optimum means that a vector was situated perpendicular or parallel to B_0. In the current case it was the vector of C4–H4 which was determined to be perpendicular to the magnetic field. This will not be true for all structures or alignments, and thus each aligned dataset was treated as above to determine if the approximated D_a value is appropriate for the data, and the value was scaled accordingly for the specific dataset if needed.

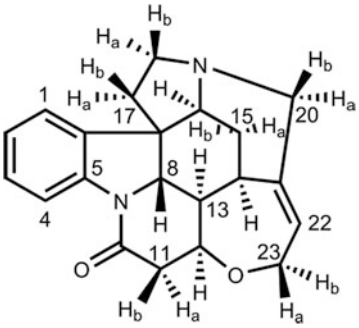
Differentiation of stereoisomers

Strychnine has previously been used for RDC calculations and multiple studies have shown that it is possible to assign the diastereotopic protons of strychnine by RDCs [9, 12, 18]. To probe whether the θ-method may be used to differentiate stereoisomers, the diastereotopic protons of strychnine were switched, fitted to the data set and compared to the right assignment. This was compared to the results of a fitting of the same data using SVD as implemented in the commercial program MSPIN. The comparison is found in Table 8.3.

The distinction was just as good for the θ-method as for SVD. The reason that the diastereotopic protons of C-15 were indistinguishable is due to very similar observed RDCs, and thus lies in the orientation of the molecules in the alignment media and not the method. Since the θ-method may easily be used in tandem with J-coupling constant and/or NOE data, the combined data was compared to 3D structures by addition of RMSDs between all experimental and theoretical data, allowing the differentiation of H-15a/H-15b, see Table 8.4. This may of course also be done for SVD fittings by separate investigation of the other parameters, but this approach is slightly more cumbersome.

Table 8.3 Comparison of the distinction of right and wrong assignment of the diastereotopic protons in strychnine by the θ-method and SVD

	Q-factor		Ratio		
Switch #	θ-method	SVD	θ-method	SVD	
Right	0.093	0.087	1.0	1.0	
Dia-11	0.688	0.565	7.4	6.5	
Dia-15	0.097	0.088	1.0	1.0	
Dia-18	0.243	0.209	2.6	2.4	
Dia-20	0.191	0.180	2.1	2.1	
Dia-23	0.523	0.458	5.6	5.3	



The diastereotopic protons are switched as stated, and may be compared to the right assignment in the top row. Aligned in PDMAA (Polymer 8.1)

Table 8.4 RMSD of RDC data, alone or coupled to NOE and/or *J*-coupling constant data for the diastereotopic protons at C15 of strychnine if assigned right or wrong

	RMSD	
	Right	Dia-15
RDC	1.247	1.266
RDC/*J*	1.853	3.207
RDC/NOE	1.287	2.247
RDC/J/NOE	1.877	4.186

Only NOEs and *J*-coupling constants involving H15a/b were used

Determining low level conformers of strychnine

A low level conformer of strychnine was previously published from NOE data, see Fig. 8.8, and recently attempts were made to use the MCST approach to determine the population of the two conformers from RDCs [8, 19]. The attempt showed varying results and as a consequence the performance of the θ-method was tested. Data from the study above as well as another dataset used in Sect. 6 were tried which may be compared to the literature [8].

In line with the MCST SVD fittings in the literature, the studies here also concluded that it was hard to obtain a meaningful conformational population for the structures of strychnine by the first set of RDCs. The two methods gave more or less exactly the same populations from the first dataset, using MCST and the θ-method alike, with an overestimation of the minor conformer as given in Table 8.5. For the second dataset the population is the same when using the θ-method while the MCST performs significantly better. Compared to the literature study by Thiele et al. the populations were in good agreement as the population of conformer 1 was populated from around 0.8 to 1 for multiple literature datasets [8].

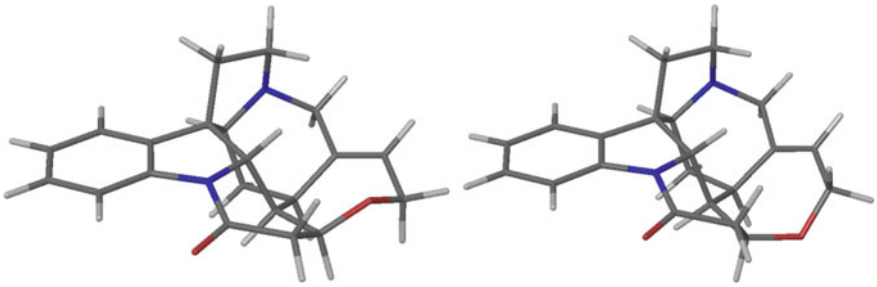


Fig. 8.8 The major (left) and minor (right) conformer of strychnine. It is noted that a third conformer has been presented, but this conformer is present well below one percent and not used for comparison. [8, 19, 20]

Table 8.5 Populations of the two conformers of strychnine, using only one bond RDCs or including long-range RDCs

	$^1D_{CH}$			$^1D_{CH}$ + $^nD_{HH}$		
	Conf. 1	Conf. 2	*Q*-factor	Conf. 1	Conf. 2	*Q*-factor
Dataset 1						
SVD	0.79	0.21	0.081	1.00	0.00	0.223
SVD, scaling	0.81	0.19	0.080	0.96	0.04	0.168
θ-method	0.81	0.19	0.072	0.88	0.12	0.164
Dataset 2						
SVD	0.96	0.04	0.079	0.98	0.02	0.296
SVD, scaling	0.96	0.04	0.079	1.00	0.00	0.119
θ-method	0.84	0.16	0.073	0.95	0.05	0.118
NOE [19]	0.98	0.02		0.98	0.02	

Scaled SVD covers the usage of "SVD Hz scaling" option in MSPIN, where the larger RDCs contribute more to the fit. This was needed to obtain good *Q*-factors when using the full datasets. (Dataset 1: Polymer 8.1, dataset 2: Polymer 6.2)

To test whether the population fit would be improved by utilizing the additional vectors from the S^3 HMBC *homo*, see Chap. 6, back-calculations were performed which included these data. The reasoning behind a possible better fit was that the additional data included internuclear vectors involving the protons 23a/23b that differ in length between the conformations. Due to the r^{-3} dependence of the resulting RDC this was thought to greatly influence the resulting populations from the data. Indeed, the inclusion of long-range RDC data generally improved the populations for both methods, as seen in Table 8.5, compared to the expected population.

The MCST approach seemed to be slightly better for this simple system. This could be attributed to the fitting procedure, where MCST uses iterative population fitting, while the θ-method is forced to rely on an additive best fit approach due to the amount of "structures" generated from the rotations. It is also beneficial for MCST that the changes in the overall structure were minor between the two

conformations, and thus the overlay of the structures was easily determined. In conclusion both methods seemed to be inferior to an NOE approach, which was not surprising due to the r^{-6} dependence leading to a huge impact by small changes in distances on the NOE intensities, while also minimizing contributions from errors.

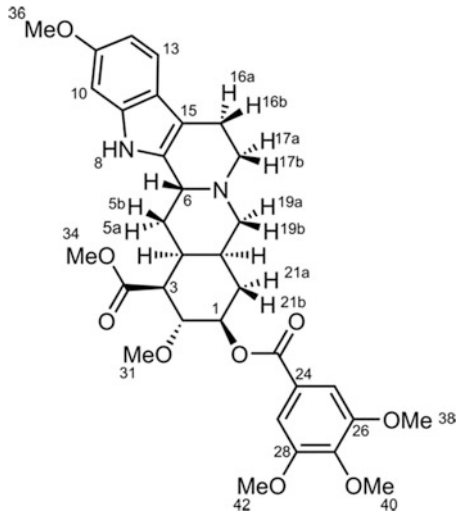
To test the scope of the θ-method other organic compounds were tested by fitting of experimental RDCs to multiple conformers. These will be presented below.

8.6.2 Reserpine

Reserpine is a plant alkaloid which features a linear pentacyclic system, see figure in Table 8.6. At least one strained ring conformation is present at all time, which was thought to lead to flexibility by differences in ring conformation between conformers. Unlike strychnine, where the molecular structure is very rigid and very few structures were identified, reserpine had have many different conformations in the simulation, though a single conformer of the ring system was identified which is populated approximately 90% from force field energies. The evaluation of many conformers could be problematic in an SVD fitting in MSPIN as the computational time increases rapidly with the number of structures. One approach could be to determine the relative energies from more accurate DFT calculations, or trust those of the force field method and remove conformers below a certain threshold, and thereby limit the amount of structures. In Fig. 8.9 the conformational space of reserpine is illustrated, with focus on the pentacyclic system.

All conformers were included in the fitting procedure disregarding any energy differences determined in the FF minimization. This should test the prowess of the θ-method and see if the correct conformer was obtained. The correlation obtained from the θ-method, and those using SVD, between experimental and back-calculated data, is found in Table 8.6. For the first SVD fitting, a single structure, representing the pentacyclic system with the lowest energy was used. For the second, structures that represented the conformational averaging of the pentacyclic system were used, due to the amount of structures else needing fitting. The θ-method utilized structures covering a much larger conformational space (114 conformers). Both methods identified the correct major conformer, but some observations indicated that the SVD method was problematic: The correlation to experimental data was markedly worse and the D_a for the multiple-conformers fitting was much higher than in the two other cases. This may be contributed to the overlay of the structures, where the heavy atoms in the pentacyclic system were used. It was not possible to establish a better overlay, so this may not be the problem. Also the limited number of input structures could have been problematic. Still, the ability to input all conformations in the θ-method, rendered all post-simulation pre-fitting considerations, such as overlay of atoms, unnecessary, as a big dataset with many structures could easily and quickly be fitted to experimental data without prior 3D structural assumptions.

Table 8.6 Experimental and back-calculated RDCs for reserpine in Hz

C	H	Exp	θ-method	SVD_{single}	SVD_{mult}	
1	1	−36.3	−36.1	−33.4	−35.2	
2	2	−35.5	−36.3	−35.7	−37.1	
3	3	−38.5	−37.5	−30.7	−24.3	
4	4	12.5	14.5	13.9	16.1	
5	5a	−18.9	−20.4	−22.9	−18.2	
5	5b	22.5	20.2	13.7	20.7	
6	6	−28.5	−28.9	−27.3	−29.2	
10	10	2.3	2.1	2.9	2.8	
12	12	27.4	29.7	30.6	40.4	
13	13	2.4	1.7	2.5	2.7	
16	16a	−55.7	−57.0	−41.3	−56.0	
16	16b	36.9	33.6	31.5	23.7	
17	17a/17b	−37.6	−31.0	–	–	
19	19a	15.8	15.4	12.5	6.5	
19	19b	−17.6	−17.5	−22.0	−20.2	
20	20	−33.4	−33.2	−30.8	−35.7	
21	21a	−27.5	−27.1	−26.5	−23.8	
21	21b	−22.9	−23.4	−24.0	−16.0	
25/29	25/29	−16.8	−15.7	–	–	
31	31	0.4	0.5	8.9	−7.4	
34	34	7.4	7.6	6.0	0.9	
36	36	−1.5	−1.8	−0.1	−12.2	
40	40	−2.0	−3.4	−1.9	−3.9	
		Q-factor	0.055	0.165	0.259	
	Q (rigid)		0.041	0.144	0.175	
		D_a	1.04E-03	1.02E-03	2.18E-03	

Calculated by SVD in MSPIN and the θ-method. The data from SVD was compared to a single structure or conformers representing the core pentacyclic system and 3-jump models of the rotatable groups. D_a is scaled by 1.3 in the θ-method. (Polymer 8.2). Right: Structure of reserpine

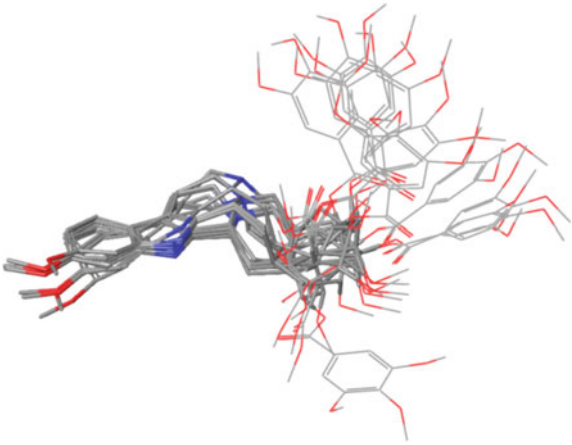


Fig. 8.9 Conformational space covered by the pentacyclic ring of reserpine

The extracted RDCs were more than adequate to discern the low from the high energy structures using both methods, and almost only structures representing the lowest relative energy of the ring system were present. The structural average correlated much better to experimental *J*-coupling constants and NOE correlations than did an average over all the structures. The major conformer from the RDC, NOE and *J*-coupling constant data parameters are found in Fig. 8.10, where the pentacyclic system is clearly identical. The conformational populations of the pentacyclic rings are visualized in Fig. 8.11 by the three dihedral angles given, as it was assumed that the indole system is rigid.

The θ-method differentiated all diastereomers and most diastereotopic protons, other than 17 due to spectral overlap, and 21 due to similar RDCs, as seen in Table 8.7. This feat was performed including quite different conformations from all diastereomers, which could all be used to favor a wrong diastereomer. This easily demonstrates the ability to find a meaningful conformational average from multiple possible structures utilizing the θ-method and solving complex 3D structural problems such as the determination of stereochemistry. The same calculations were performed using NOE and *J*-coupling constant data (Tables 8.8 and 8.9). The data complemented each other well, and the RDCs were just as discriminative as the other data. In the case of the diastereotopic protons at C-16, RDCs were actually needed in order to discern the two. This was caused by the overlap of the resonances of H17a and H17b which led to no discerning NOE correlations of H-16a/b, and also no *J*-coupling constants involving H-16a/b were determined. If all of the data types are utilized, the stereochemistry of all chiral and pro-chiral centers was easily solved (Table 8.10).

The methoxy-groups could rotate and were all implicitly averaged in the fitting of the data, both using the θ-method and SVD fitting to a single tensor in MSPIN. Information regarding the methoxy ester could be discarded as only minor differences in back-calculated RDCs in a 3-jump model were observed due to similar

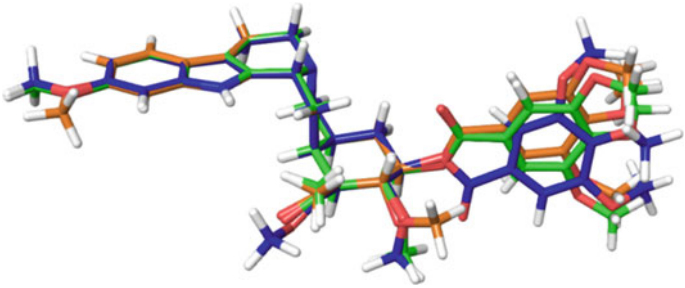


Fig. 8.10 The 3D structure which best fit the NMR data. Green: RDC data, blue: *J*-coupling constant data (RMSD = 0.024 compared to RDC) and orange NOE data (RMSD= 0.013 compared to RDC). The RMSD is based on the heavy atoms in the pentacyclic system only

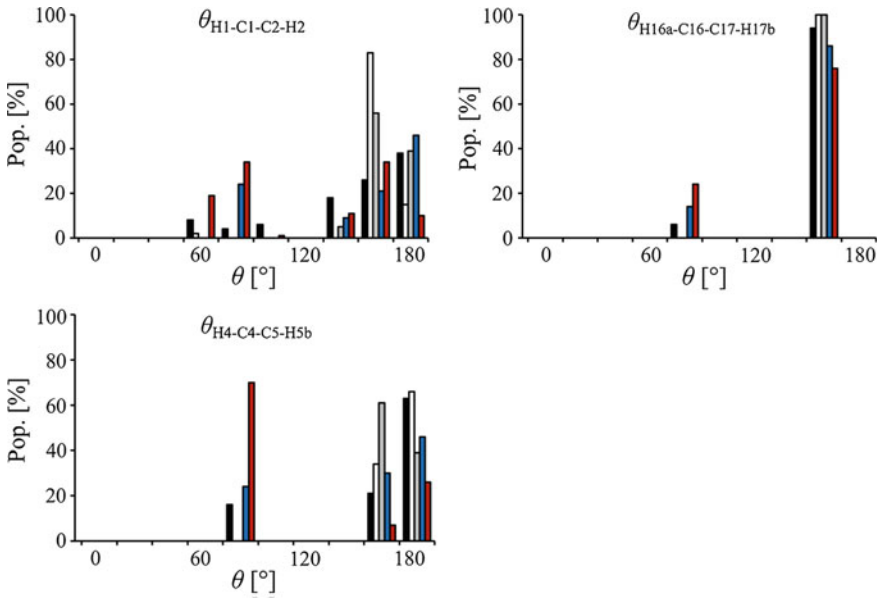


Fig. 8.11 Comparison of dihedral angles representing the ring-conformation of the pentacyclic system as found from θ-method (black), NOEs (white), *J*-coupling constants (grey), SVD (blue) and all possible structures (red). Groups cover $\pm 20°$ and no populations were found below 0°. All structures in a 50 kJ/mol window were included in the fit as equally probable structures

angles to the magnetic field, as the ester bond was static. The phenol alcohol could be described as a two jump model. The population percentages of each position are found in Table 8.11 and are in good correlation with NOE populations.

A 3-jump model was assumed for the methoxy group C31 and the populations are seen in Table 8.12. While there are discrepancies between the RDC and NOE data, the RDC data consisted of a single internuclear vector, while three NOE

Table 8.7 RDC: *Q*-factors and ratios of *Q*-factors for wrongly assigned diastereomers of reserpine versus right assignment

Switch #	*Q*-factor	Ratio
Right	0.071	1.0
Dia-1	**0.630**	**8.9**
Dia-2	**0.518**	**7.3**
Dia-3	0.129	1.8
Dia-4	**0.574**	**8.1**
Dia-6	**0.709**	**10.0**
Dia-20	**0.768**	**10.8**
Dia-5	0.188	2.6
Dia-16	**0.457**	**6.4**
Dia-19	0.108	1.5
Dia-21	0.070	1.0

Calculated by θ method. 17 not included due to spectral overlap of protons. Bold: easily distinguished

Table 8.8 NOE: MAE% and ratios of MAE% for wrongly assigned diastereomers of reserpine versus right assignment from NOE (ISPA)

Switch #	MAE%	Ratio
Right	5.6	1.0
Dia-1	**22.3**	4.0
Dia-2	9.3	1.7
Dia-3	10.1	1.8
Dia-4	**12.8**	**2.3**
Dia-6	**14.0**	**2.5**
Dia-20	**18.5**	**3.3**
Dia-5	7.1	1.3
Dia-16	7.7	1.4
Dia-19	8.5	1.5
Dia-21	10.2	1.8

Bold: easily distinguished

correlations were involved in the population determination. The NOE data was thus deemed more trustworthy.

It should be stressed that the θ-method, while slower for single compounds compared to SVD calculations, performed much faster when optimizing fits of multiple structures to experimental data where the SVD optimization procedure is slow. The ability to quickly screen multiple conformers and identify the correct ones is important, and while it is in principle possible to decrease calculation times for SVD methods by screening a lot of conformers in sequence, it is cumbersome, and does not really evaluate a possible fit of multiple structures. The θ-method may here be employed to quickly select the structures which fit experimental data.

Table 8.9 *J*-coupling constants: MAE and ratios of MAE for wrongly assigned diastereomers of reserpine versus right assignment

Switch #	MAE	Ratio
Right	1.0	1.0
Dia-1	**6.0**	**6.0**
Dia-2	2.3	2.3
Dia-3	2.0	2.0
Dia-4	**3.8**	**3.8**
Dia-6	1.3	1.3
Dia-20	**4.8**	**4.8**
Dia-5	1.7	1.7
Dia-16	–	–
Dia-19	1.0	1.0
Dia-21	2.4	2.4

Calculated by HLA. Bold: easily distinguished

Table 8.10 Including NOEs, *J*s and RDCs in the fitting of wrongly assigned diastereomers of reserpine versus right assignment

Switch	NOE	*J*	RDC	Ratio				
#	MAE%	MAE%	Q	Σ	NOE	*J*	RDC	Σ
Right	6.0	15.7	0.112	0.38	1.0	1.0	1.0	1.0
Dia-1	**24.8**	**55.4**	**0.636**	**1.77**	**4.1**	**3.5**	**5.7**	**4.7**
Dia-2	**9.4**	**30.4**	**0.533**	**1.03**	**1.6**	**1.9**	**4.8**	**2.7**
Dia-3	10.1	28.8	0.136	0.61	1.7	1.8	1.2	1.6
Dia-4	**12.8**	**49.8**	**0.594**	**1.22**	**2.1**	**3.2**	**5.3**	**3.2**
Dia-6	**10.0**	**18.9**	**0.544**	**0.96**	**1.7**	**1.2**	**4.9**	**2.5**
Dia-20	**18.6**	**53.5**	**0.764**	**1.68**	**3.1**	**3.4**	**6.8**	**4.4**
Dia-5	7.6	22.9	0.264	0.66	1.3	1.5	2.4	1.7
Dia-16	**6.6**	**15.1**	**0.545**	**0.84**	**1.1**	**1.0**	**4.9**	**2.2**
Dia-19	8.9	14.6	0.185	0.54	1.5	0.9	1.7	1.4
Dia-21	**18.2**	**28.0**	**0.089**	**0.75**	**3.0**	**1.8**	**0.8**	**2.0**

Calculated by ISPA, HLA or θ method. $\Sigma = MAE_{NOE} + MAE_J/10 + Q_{RDC}$. Bold: easily distinguished

Table 8.11 Orientation of the C36 methyl group as determined from RDC (θ-method, SVD) and NOE data

Conf.	θ method (%)	MSPIN (%)	NOE (%)
1	66	74	67
2	34	26	33

Table 8.12 Orientation of the C31 methyl group as determined from RDC (θ-method, SVD) and NOE data

Conf.	θ method (%)	MSPIN (%)	NOE (%)
1	54	44	85
2	8	10	10
3	38	46	5

8.6.3 Cinchona Alkaloids

The cinchona alkaloids, named from their natural source, exhibit similar structural features with multiple rotational bonds connecting rigid parts. Possibly due to their usage in e.g. organic synthesis all structures are described quite well in literature [21–24] Three alkaloids were investigated in order to test more flexible compounds: Quinine, cinchonine and cinchonidine. The conformational space of these compounds has been previously examined, and the literature thus provides reference studies [21, 22]. The cinchona alkaloid quinine was also investigated using NOEs in Sect. 3.4.1.

Cinchonidine

The alkaloid cinchonidine (Fig. 8.12) is related to quinine and the structures share molecular features only differing in a methoxy group at C-8. The RDC data were fitted to multiple structures using the θ-method, see Tables 8.13 and 8.14.

The conformations of cinchonidine were previously investigated from energy computations and NMR data, primarily via *J*-coupling constants by Bürgi et al. [21]. and Urakawa et al. [25]. These studies were based on a few optimized structures and their relative energies. The approach of this study thus differs, as the relative energy was disregarded and multiple conformers were used, but still the data is compared in Table 8.15. Since multiple conformers were used, and more dihedral angles are possible, the dihedral angles reported are centered on the angles determined from DFT studies. This may be viewed as the DFT structures being centroids.

It is noted that the SVD method, while giving a reasonable *Q*-factor did not result in a conformational population that correlated to the rest of the NMR data. The overlay was made using the heavy atoms in the aromatic system, since it exhibited the largest absolute RDCs (after correction for the fact that the values of RDCs of parallel vectors to the magnetic field are doubled compared to perpendicular vectors). The θ-method on the other hand resulted in surprisingly good populations compared to the relative energies, and compared to the NOE data of quinine, a similar structure investigated in Sect. 3.4.1 differing only by a methoxy group at C4′, the populations fit very well between the different data types.

For comparison RDCs were also obtained from DMSO. The populations in water were used, as no data in DMSO was identified in the literature, and the energies of water and acetone are almost equal while the dielectric constant of DMSO is almost just between the two. Also the *J*-coupling constant of H11-H12 is almost equal in water and DMSO, 7.5 and 7.4 Hz respectively. Theoretically, from

calculations and the *J*-coupling constant of 7.4 Hz, the ratio of $\tau 3$ dihedral angle populations should now be opposite compared to chloroform populations, which was indeed observed. The *J*-coupling constants for H11 to H12 were measured to be 3.0 Hz in chloroform and 7.4 Hz in DMSO. The back-calculated *J*-coupling constants, as the average of the *J*-coupling constants of the structures determined from the RDC data by the HLA method, gave 3.2 and 5.9 Hz respectively for chloroform and DMSO in good agreement to the theoretical values.

Cinchonine

Cinchonine was compared to the work of Kowalik et al. [22]. As for the other structures, diastereomers of internuclear vectors with numerically similar RDCs were hard to differentiate (Table 8.16). Also differentiation of the correct assignment and the diastereomer epi-cinchonine could not initially be done based in the RDC data, Table 8.17. It should be noted that this was due to the choice of using all structures with no regard to their simulated energy, which was in this case not a preferable approach. If a cutoff was chosen at below 30 kJ/mol, it was possible to discern the stereoisomers, and it is in principle possible to solve the stereochemistry from RDCs and knowledge of relative energies between conformers.

Another approach is to utilize the *J*-coupling constant between H11 and H12. For epi-cinchonine this *J*-coupling constant is reported as 9.9 Hz in chloroform compared to 4.0 Hz for cinchonine, and the diastereomers could be distinguished from the conformational differences and the associated value of the average coupling constant, where the wrong assignment (epi-cinchonine) utilized conformations which led to a markedly different *J*-coupling constant, Table 8.16 [22]. Populations determined of the two diastereomers from the cinchonine RDC data are found in Table 8.18.

It was possible to fit the multiple conformers to the data using SVD to give a reasonable result, but there was a major concern. The alignment strength was much higher using SVD, which is rather problematic as this correlates to the individual conformers having theoretical RDCs in the hundreds of Hz, which then averages to give reasonable back-calculated RDCs. This does not seem reasonable but the cause of this was not definitely determined. It is speculated that this was caused by the rigidity in the alignment as introduced by the MCST methodology. Since all structures are overlaid, the pre-fitted conformers are not allowed to average the RDCs individually and an overall average is constructed instead. In that regard it should be noted that the tensor is averaged along a fixed axis and it may well be that the axis differs between individual structures, leading to bad fit of the system. The problem was not alleviated by overlaying another part of the structure.

One could maybe limit the possible alignment strength determined by the SVD fitting, but the author has no suggestions as to how, as any limitation will result in a worse fit between experimental and back-calculated values in a natively best fit method. It was tried to utilize the alignment tensors of the individual structures as a base, by an in house written Matlab® script. Unfortunately this always resulted in the structure by which the alignment tensor was established, being by far the best

Table 8.13 Experimental versus back-calculated RDCs for cinchonidine

C	H	Exp (Hz)	θ-method	SVD
2	2a	3.9	4.7	5.8
2	2b	10.8	8.3	5.3
3	3	24.5	23.2	24.1
4	4	27.9	23.1	22.0
5	5a	−18.5	−21.2	−20.9
5	5b	29.7	25.8	26.0
6	6a	−38.6	−37.7	−39.1
6	6b	5.0	9.0	8.8
7	7a	−21.0	−20.3	−22.4
7	7b	−22.7	−22.3	−22.0
8	8	−5.3	−7.0	−9.2
9	9	15.8	16.1	18.0
10	10	26.2	23.9	21.0
11	11a	−27.0	−28.8	−23.4
11	11b	21.3	24.4	21.1
2′	2′	22.3	24.3	23.5
3′	3′	23.5	22.3	22.3
5′	5′	27.8	25.2	28.1
6′	6′	25.1	23.0	23.1
7′	7′	20.9	21.5	22.6
8′	8′	25.6	25.1	26.5
		Q-factor	0.100	0.127
		D_a	8.42e-04	1.06e-3

Calculated by θ-method or SVD. (Polymer 8.3)

Table 8.14 *Q*-factors and ratios of *Q*-factors for wrongly assigned diastereomers of cinchonidine versus the right assignment

Switch #	*Q-factor*	Ratio
Right	0.100	1.0
Dia-9	0.331	3.1
Dia-7	0.102	1.0
Dia-2	0.106	1.0
Dia-5	0.372	3.5
Dia-6	0.410	3.8
Dia-11	0.497	4.6

Calculated by the θ-method

fit. Also linear combinations of the alignment tensors were tried, which always resulted in worse fit compared to the native alignment tensors, and there is no theoretical evidence that this should be a viable approach.

The populations from the θ method correlated less well to the theoretical populations than those of cinchonidine, but still the fit to the theoretical populations

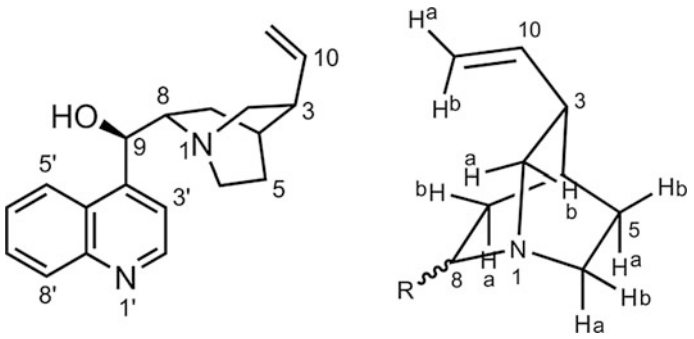


Fig. 8.12 Structure of cinchonidine and the orientation of diastereotopic protons for all cinchona alkaloids included in this thesis

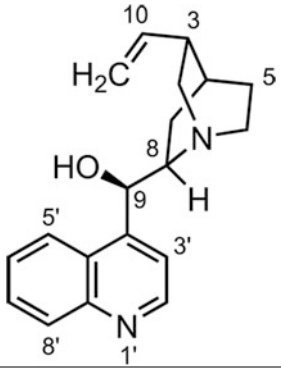
Table 8.15 Conformational populations which exhibit dihedrals τ, defined as the value $\pm 20°$, in percent (%)

	$\tau 1$			$\tau 2$				$\tau 3$					
	80	−100	40	150	−50	−90	50	−80	70	40	180	Q	D_a
$CDCl_3$													
Ref, energy (1)	87	13	–	74	–	–	26	74	–		26	–	–
Ref, energy (2.1)	80	11	9	69	–	7	24	69	–	7	24	–	–
Ref, energy (2.2)	89	6	5	77	–	12	11	77	–	12	11	–	–
θ method, no scaling	48	26	26	86	–	–	14	86	–		14	0.168	6.48e-4
θ method, scaled by 1.3	58	14	28	75	–	–	25	75	–		25	0.100	8.42e-4
SVD	55	45	–	100	–	–	–	100	–	–	–	0.159	1.06e-3
NOE (p. 57, quinine)	67	12	21	67	–	–	33	67	–	–	33	–	–
DMSO-d_6													
Ref, energy (water)[a]	72	27	–	45	–	–	54	45	–	–	54	–	–
θ method, scaled by 1.2	39	46	15	28	20	–	52	28	20	–	52	0.052	3.96e-4

$\tau 1$: C3–C4–C9–C8, $\tau 2$: C4–C9–C8–N, $\tau 3$: H9–C8–C9–H8. Population fitting used in MSPIN. The population of an NOE fit is used in comparison, using that of quinine, which has a similar structure as discussed in Sect. 3.4.1. *Other conformations identified, does not sum to 100.[a]See text. ($CDCl_3$: Polymer 8.3, DMSO-d_6: Polymer 8.4)

Table 8.16 Q-factors and ratios of *Q*-factors for wrongly assigned dia-stereomers of cinchonine versus right assignment

Switch #	*Q*-factor	Ratio
Right	0.096	1.0
Dia-9	0.095	1.0
Dia-7	0.098	1.0
Dia-2	0.373	3.9
Dia-6	0.167	1.7
Dia-5	0.099	1.0



Calculated by the θ method

was reasonably good, also considering the differences observed between energy calculations and NOE conformer populations for quinine.

Quinine

The final cinchona alkaloid to be investigated by RDCs using the θ-method is quinine (Fig. 8.13). The conformational space of quinine was already discussed in Sect. 3.4.1 obtained using NOE distances. As DFT optimized conformers of quinine were already obtained, quinine was chosen to determine the effect of having either a few, highly optimized structures (centroids) or multiple structures (Table 8.19).

Fig. 8.13 Structure of quinine

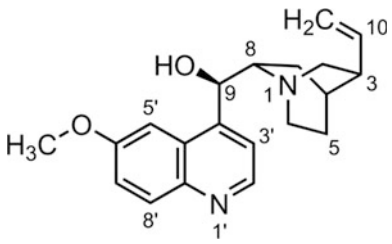


Table 8.17 Experimental versus back-calculated RDCs for cinchonine and the wrong diastereoisomer epi-cinchonine

C	H	Exp (Hz)	θ method			
			Cinchonine (50 kJ/mol)	Cinchonine (30 kJ/mol)	Epi-cinch. (50 kJ/mol)	Epi-cinch. (30 kJ/mol)
2	2a	−33.1	−33.4	−33.4	−33.3	−36.6
2	2b	9.0	7.6	7.6	8.1	13.8
3	3	−1.5	−0.7	−0.7	−0.3	1.6
4	4	3.7	6.9	6.9	7.5	2.1
5	5a	−5.3	−3.0	−3.0	−3.4	−1.3
5	5b	−3.6	−1.2	−1.2	−1.5	−0.2
6	6a	19.2	21.1	21.1	20.2	21.5
6	6b	8.1	8.6	8.6	8.8	14.7
7	7a	−10.6	−7.5	−7.5	−7.3	−6.1
7	7b	−17.0	−13.1	−13.1	−12.9	−9.2
8	8	3.2	0.9	0.9	1.9	7.2
9	9	13.8	13.9	13.9	13.8	6.8
10	10	−7.6	−7.3	−7.3	−7.6	−7.6
2′	2′	30.1	29.8	29.8	29.9	27.7
3′	3′	24.7	24.8	24.8	24.6	20.7
5′	5′	23.9	24.6	24.6	24.7	17.6
6′	6′	29.5	29.5	29.5	29.3	28.4
7′	7′	23.0	23.4	23.4	23.0	19.0
8′	8′	24.8	24.2	24.2	24.0	17.5
		Q-factor	0.096	0.097	0.095	0.250
		D_a	8.30E-04	8.30E-04	8.30E-04	8.30E-04
Resulting *J*			4.8 Hz		2.7 Hz	

Calculated by θ method, D_a scaled by 1.3. *J*-coupling constants calculated using the HLA equation and the conformer populations from the RDC fit. (Polymer 8.5)

The RDC data was fitted to the seven structures that were used to describe the conformational space of quinine for NOEs in Sect. 3.4.1.

The populations are seen in Table 8.20, comparing to theoretical conformer populations and those from the θ-method and SVD.

Analogues to the SVD fitting of cinchonine, the degree of alignment for quinine was much too high when using the aromatic system to overlay the structures, and a bit better when using the bridged system. For both the θ method and SVD, the general populations were acceptable for the rotation of the aromatic system, while the rotation around the C9–C8 bonds was not populated as expected, when compared to energy-or NOE derived populations. It was thus tried to include more structures, departing the centroids, in the approach that worked well for the cinchona alkaloids previously investigated. A reason for a worse population fit could be that the dataset was the smallest of the investigated cinchona alkaloids, due to

Table 8.18 Conformational populations which exhibit dihedrals τ, defined as the value $\pm 20°$, in percent (%)

	$\tau 1$			$\tau 2$				
Cinchonine	40	−135	80	55	−60	160	Q	D_a
Ref, energy [22]	95	5	–	19	74	6	–	–
θ method, scaled by 1.6	54	33	13	34	52	7*	0.096	8.304E-04
SVD	39	48	13	47	30	12*	0.095	4.761E-03[a]
	$\tau 1$			$\tau 2$				
Epi-cinchonine	−35	130	−110	55	−65	175		
Ref, energy [22]	51	49	–	–	32	68	–	–
θ method, scaled by 1.6	54	15	31	12	73	15	0.095	8.300E-04

$\tau 1$: C3–C4–C9–O, $\tau 2$: O–C9–C8–C7. Population fitting used in MSPIN. *Other conformations identified, does not sum to 100. [a]SVD fitting holds little meaning due to high D_a, see text. (Polymer 8.5)

structural overlaps and being acquired at 400 MHz, which might be problematic in the population fitting.

Entering more structures from force field calculations resulted in populations that were more alike to that of cinchonine and cinchonidine, where populations given by the rotation around C2–C9 did not correlate well to energies, but okay to populations from NOE distances, while the conformer population given by the rotation around the C8–C9 bond fitted reasonably well to the theoretical and NOE data, see Table 8.21. Thus the inclusion of multiple structures from force field simulations led to populations which correlate much better to energy- and NOE derived populations, largely comparable to the populations from the data of cinchonidine.

It was difficult to pinpoint the rather large differences between populations identified using few and many conformers, other than the fact that multiple conformers may better illustrate the actual structural space. Still it markedly limits the usage of the θ-method, and more data is needed to test reproducibility of the RDC data.

Generally the populations identified from RDC data for the cinchona alkaloids correlate nicely to energies and, especially, other NMR observables. This was actually rather surprising and the results are promising, since the cinchona alkaloids are more flexible than the structures usually investigated by RDCs, and it actually may be that RDCs hold useful non-redundant information on population of such systems.

8.6.4 8-Phenyl-Methol

The possible pitfall of utilizing an SVD based mulita-conformer analysis, in the form of overfitting, should be addressed. For this a literature study is re-calculated by the θ-method, to investigate assumptions of overfitting.

Table 8.19 Experimental versus calculated RDCs for quinine

C	H	Exp (Hz)	θ method		SVD	
			MM	DFT	SVD[a]	SVD[b]
3	3	12.3	12.6	12.7	12.8	12.7
4	4	9.6	10.5	11.3	9.8	11.3
6	6a	−26.5	−25.5	−25.2	−24.8	−25.2
6	6b	2.0	3.2	3.3	2.8	3.3
8	8	−2.8	−2.0	−1.3	−1.7	−1.3
9	9	10.1	10.2	9.9	10.4	9.9
2’	2’	16.3	16.6	16.4	16.6	16.4
3’	3’	15.5	14.9	14.1	14.7	14.1
5’	5’	14.6	14.3	14.3	16.9	14.3
7’	7’	13.2	13.7	14.5	14.4	14.5
8’	8’	13.5	13.7	13.6	11.9	13.6
		Q-factor	0.048	0.071	0.079	0.107
		D_a	5.54E-04	5.45E-04	4.50E-03	1.19E-03

Calculated by θ-method using structures from MM or DFT optimized structures. D_a was scaled by 1.6. For SVD only DFT optimized structures were used, overlaid by the aromatic (a) or the bridged (b) part. (Polymer 8.6)

The structural conformation of 8-phenylmenthol was previously determined [26]. Here the assignment of the pro-chiral methyl groups was explored by RDCs; see Table 8.22 for the structure. Three conformers were identified by DFT calculations, which are representing a 3-jump model, i.e. conformers with the phenyl-group at all three positions indicated in Table 8.22. The RDC data were fitted to the conformers, with no assumptions of populations, in a least squares sense, from the values calculated by SVD from a common alignment tensor. The result is surprising in that the populations are almost equal and thus unaffected by the assignment of the methyl groups. This finding leads to one assignment being a better fit to the RDC data and the authors use this finding to speculate that it may be used in differentiation of the assignment. The fit between populations and RDCs seems puzzling to the current author as it is non-obvious that the population should be equal between assignments, and an identical fit should be possible from both pro-chiral assignments by inversion of the populations. It should be noted that the conclusions of the publication is not disputed, as other data is used in support.

For the investigation of the data by the θ-method, the published RDC data was compared to back-calculated values obtained from the published conformers as well as eight conformations covering a 3-jump model with slight variations in the rotation. The phenyl group was treated as a 2-jump model, in practice averaging the RDCs of *ortho*- and *meta*-^{13}C-^{1}H vectors in accordance with the original publication. A scaling factor of 1.2 was used for D_a, positioning the ring firmly in space. The results are summarized in Table 8.22.

Table 8.20 Conformational populations which exhibit dihedrals τ, defined as the value $\pm 20°$, in percent (%)

	$\tau 1$			$\tau 2$			$\tau 3$				
	100	−100	20	150	−70	50	−80	60	180	Q	D_a
Urakawa, energy [25]	79	11	10	72	3	25	72	3	25		
Butts, energy (see Sect. 3.4.1)	75	13	12	79	0	21	79	0	21		
θ-method, no scaling	78	22	0	56	0	44	56	0	44	0.138	3.405e-4
θ-method scaled by 1.6	59	31	0	31	7	62	31	7	62	0.071	5.448e-4
SVD (arom)[a]	34	51	15	29	37	34	29	37	34	0.079	4.50E-03
SVD (bridge)	75	9	16	42	0	58	47	0	58	0.107	1.19E-03
NOE (p.57)	67	12	21	67	0	33	67	0	33		

$\tau 1$: C3-C4-C9-C8, $\tau 2$: C4-C9-C8-N, $\tau 3$: H9-C8-C9-H8. D_a was scaled by 1.5. Population fitting used in MSPIN. [a]SVD fitting holds little meaning due to high D_a, see text. (Polymer 8.6)

Table 8.21 Conformational populations which exhibit dihedrals τ, defined as the value $\pm 20°$, in percent (%)

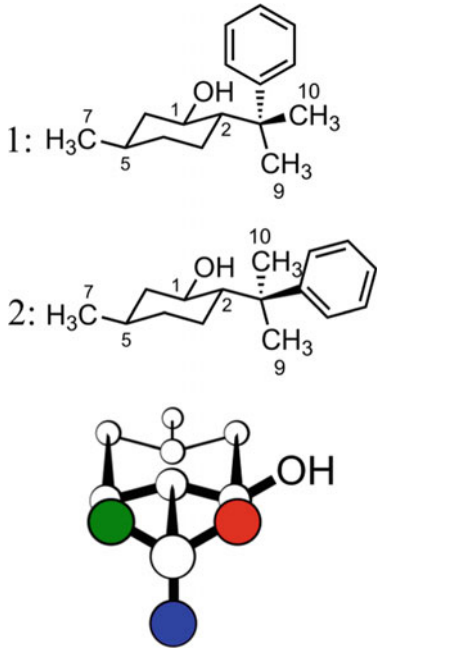
	$\tau 1$			$\tau 2$			$\tau 3$				
	90	−100	20	150	−80	50	−80	60	180	Q	D_a
Urakawa, energy [25]	79	11	10	72	3	25	72	3	25		
Butts, energy[III]	75	13	12	79	0	21	79	0	21		
θ-method, no scaling	43	16	41	91	–	9	91	–	9	0.144	3.46e-4
θ-method scaled by 1.6	55	18	27	55	17	27*	55	17	27*	0.048	5.54e-4
NOE (p. 57)	67	12	21	67	–	33	67	–	33	–	–

$\tau 1$: C3–C4–C9–C8, $\tau 2$: C4–C9–C8–N, $\tau 3$: H9–C8–C9–H8. *Other conformations identified, does not sum to 100. (Polymer 8.6)

When fitting the conformations with the θ-method excellent fits of experimental and back-calculated RDCs for both methyl assignments were determined. The assumption that the conformational population of G+ and G− could be inverted to fit the data between methyl assignments was confirmed. This makes it impossible to differentiate the two pro-chiral methyl groups, in good correlation to the experimental data. Unless one can unambiguously establish the alignment of the chair, the free rotation of the group should make the data a possible match for both assignments. Also, it should be noted that since the chair only has four non-parallel vectors, the rotation of the group is used in establishing the alignment tensor, and thus relative alignment of the two groups is hardly established.

Table 8.22 Experimental [26] versus calculated RDCs for 8-phenyl-methol

C	H	Exp (Hz)	8 structures (Hz)		3 structures (Hz) [26]	
			Stereo 1	Stereo 2	Stereo 1	Stereo 2
1	1	113.5	111.8	111.0	110.9	110.2
2	2	114.3	114.3	113.4	112.9	114.8
4	4 eq	7.7	5.9	6.6	8.2	7.4
4	4ax	112.0	113.6	112.0	113.1	111.9
5	5	103.8	103.5	103.4	106.4	105.2
6	6 eq	8.1	6.0	7.2	7.5	7.6
6	6ax	112.4	111.4	114.1	110.8	112.5
7	7	33.9	38.0	38.8	37.9	37.3
9	9	−32.6	−32.7	−32.2	−30.7	−31.5
10	10	−13.0	−13.3	−13.6	−12.0	−13.1
ortho	*ortho*	105.3	100.2	101.1	103.8	104.2
meta	*meta*	102.4	105.1	105.5	101.1	101.7
para	*para*	−16.9	−16.2	−16.2	−16.5	−16.9
		Q-factor	0.021	0.021	0.023	0.018
		D_a	3.76E-03	3.76E-03	3.01E-03	3.01E-03
Populations						
		Anti	0.03	0.00	0.00	0.00
		G+	0.16	0.62	0.20	0.62
		G−	0.81	0.38	0.80	0.38



Calculated by θ-method. The orientation of the methyl groups 9 and 10 are interchanged between stereoisomers as indicated to the right. Eight structures centered on a 3-jump model for the rotation around the C2–C8 bond or three structures previously reported were used in the fit. [26] The rotation of the C2–C8 bond is named given by the orientation of the phenyl group: Anti (blue), G+ (green) and G− (red)

This is a prime example of how MCST fitting will at times lead to the overfitting of data. It should be noted that the authors offer another likely method of differentiation; by chemical shift calculations as the aromatic ring will shift resonances due to the inductive effect [26]. This makes the G− conformer the most likely highly populated of the G+ and G− conformers, and thus the assignment in structure 1 (Table 8.22) is the most likely assignment, in line with the conclusions reached in the original publication. Thus coupling of different data may be used to differentiate the two pro-chiral methyl groups; RDCs alone may not.

8.7 The Combination of Multiple Methods

The three NMR observables primarily utilized in this thesis, NOEs, ^3J-coupling constants and RDCs, are highly compatible and orthogonal methods to gain information of 3D structures, due to differences in the mechanisms that give rise to the observables and in the averaging functions. For small molecules the three observables are seldom combined and most literature is centered on either NOEs and *J*-coupling constants or RDCs. A reason for this may be that RDCs are not yet widespread in the small molecule community and may be seen as cumbersome to obtain. Another might be that multiple programs are used for the respective calculations that do not necessarily interface well. Also the averaging of structures for RDC calculations, as well as NOEs and *J*-coupling constant calculations to a large extent, is not always straightforward if large numbers of structures are obtained.

The θ-method utilizes RDC data so that it may be easily coupled to data from NOEs and *J*-coupling constants, and it is expected that this may lead to an increase of structural knowledge and ease of calculations. The three data types have been programmed to be compared to, and converge with, each other, in order to investigate 3D structures of organic compounds. This is done through the ISPA method, Karplus equations (including HLA) and the θ-method. It is here utilized that all three methods are easily interfaced with each other, for single or multiple conformer systems alike.

8.8 Conclusion and Perspectives

Population fitting of flexible compounds from RDCs still lies in the future. More data from a larger number of structures is needed to determine the proper method for structural averaging in RDC calculations. The θ-method is a possible solution, where the spectroscopist or (computational) chemist is relieved from making decisions of overlays of structures and thus effectively parts of the molecule that are situated equally in space across conformers, thus easing the work. Whether the assumptions that D_a may be determined from the experimental data and that a finite rotation will lead to a good representation of data holds true, may need to be finally

determined from more data, but the present results are very promising. The θ-method discerned stereoisomers just as well as SVD, for the relatively rigid compounds investigated. The populations from the θ-method for the cinchona alkaloids correlated well to theoretical energy calculations as well as NOE populations. This is actually a little surprising but it seems to confirm an approximation made initially; that the conformer population is independent of the alignment media and D_a is identical for all conformers. Whether this hold true for more compounds will need to be investigated in future work.

It is readily realized that more flexibility is needed in small molecule RDC calculations, if the technique is to find widespread usage. Otherwise the more easily accessible observables will be used instead, and for the very rigid structures that are routinely used in RDC publications, even chemical shift calculation may be an equally good and much easier alternative. This is not the case regarding flexible structures, where NOEs, *J*-coupling constants and RDCs are needed for structural information. It is thus imperative that methods are developed which may cope for structural flexibility, and this is especially true for RDCs. From most data obtained this far, which is also the data presented herein, the θ-method works well, leading to meaningful conformations as well as populations, while easily handling inputs of multiple structures.

8.9 Experimental

Polymers

See Sect. 6.4 for information regarding synthesis of the PDMAA polymer. All other necessary information is available in figures and tables throughout this chapter.

In the table below the constituents in the synthesis of the relevant polymers are given.

Name	Analyte (c [mM])	Solvent	Monomer	Crosslinker (mol%)	Monomer:solvent in synthesis (v:v)	Field (MHz)
Polymer 8.1	Strychnine (269.1)	$CDCl_3$	DMAA	MBAA (0.34)	–	800
Polymer 8.2	Reserpine (132.4)	$CDCl_3$	DMAA	MBAA (0.17)	1:1 $CDCl_3$	800
Polymer 8.3	Cinchonidine (277.1)	$CDCl_3$	DMAA	MBAA (0.17)	1:1 $CDCl_3$	800
Polymer 8.4	Cinchonidine (288.7)	DMSO-d_6	DMAA	MBAA (0.16)	5:1 $CDCl_3$	400
Polymer 8.5	Cinchonine (219.4)	$CDCl_3$	DMAA	MBAA (0.17)	1:1 $CDCl_3$	600

(continued)

(continued)

Name	Analyte (c [mM])	Solvent	Monomer	Crosslinker (mol%)	Monomer:solvent in synthesis (v:v)	Field (MHz)
Polymer 8.6	Quinine (219.5)	$CDCl_3$	DMAA	MBAA (0.17)	1:1 $CDCl_3$	400

NMR experiments

The CLIP-HSQC experiment was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. Isotropic spectra of strychnine (132 mM), reserpine (53 mM), cinchonidine (77 mM), cinchonine (86 mM) and quinine (89 mM) were acquired in 5 mm tube (0.5 mL).

Simulations

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations [27], using the program MacroModel version 10.8 [28, 29]. The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization [28, 30].

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations [31]. Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

References

1. J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, J. Magn. Reson. **138**, 334–342 (1999)
2. A. Navarro-Vázquez, Magn. Reson. Chem. **50**, S73–S79 (2012)
3. A. Canales, J. Jiménez-Barbero, M. Martín-Pastor, Magn. Reson. Chem. **50**, S80–S85 (2012)
4. M. Zweckstetter, Nat. Protoc. **3**, 679–690 (2008)
5. C.M. Thiele, Concepts Magn. Reson. Part A **30A**, 65–80 (2007)
6. Matlab documentation—SVD http://se.mathworks.com/help/matlab/ref/svd.html. Accessed 15 Apr 2016
7. C.M. Thiele, Eur. J. Org. Chem. 5673–5685 (2008)
8. A. Kolmer, L.J. Edwards, I. Kuprov, C.M. Thiele, J. Magn. Reson. **261**, 101–109 (2015)
9. C.M. Thiele, S. Berger, Org. Lett. **5**, 705–708 (2003)
10. M. Ottiger, A. Bax, J. Am. Chem. Soc. **121**, 4690–4695 (1999)
11. L. Verdier, P. Sakhaii, M. Zweckstetter, C. Griesinger, J. Magn. Reson. **163**, 353–359 (2003)
12. J.D. Snider, E. Troche-Pesqueira, S.R. Woodruff, C. Gayathri, N.V. Tsarevsky, R.R. Gil, Magn. Reson. Chem. **50**, S86–S91 (2012)
13. R. Montalvao, C. Camilloni, A. De Simone, M. Vendruscolo, J. Biomol. NMR **58**, 233–238 (2014)
14. C. Camilloni, M. Vendruscolo, J. Phys. Chem. B **119**, 653–661 (2015)

15. A.O. Frank, J.C. Freudenberger, A.K. Shaytan, H. Kessler, B. Luy, Magn. Reson. Chem. **53**, 213–217 (2015)
16. G.M. Clore, A.M. Gronenborn, N. Tjandra, J. Magn. Reson. **131**, 159–162 (1998)
17. M.H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd edn. (John Wiley & Sons Ltd, Chichester, 2008)
18. C.M. Thiele, J. Org. Chem. **69**, 7403–7413 (2004)
19. C.P. Butts, C.R. Jones, J.N. Harvey, Chem. Commun. **47**, 1193–1195 (2011)
20. G. Bifulco, R. Riccio, G.E. Martin, A.V. Buevich, R.T. Williamson, Org. Lett. **15**, 654–657 (2013)
21. T. Burgi, A. Baiker, J. Am. Chem. Soc. **120**, 12920–12926 (1998)
22. J.T. Kowalik, T. Lipinska, B.J. Oleksyn, J. Sliwinski, Enantiomer **4**, 389–410 (1999)
23. G.D.H. Dijkstra, R.M. Kellogg, H. Wynberg, J. Org. Chem. **55**, 6121–6131 (1990)
24. E.M.O. Yeboah, S.O. Yeboah, G.S. Singh, Tetrahedron **67**, 1725–1762 (2011)
25. A. Urakawa, D.M. Meier, H. Rüegger, A. Baiker, J. Phys. Chem. A **112**, 7250–7255 (2008)
26. V.M. Sánchez-Pedregal, R. Santamaría-Fernández, A. Navarro-Vázquez, Org. Lett. **11**, 1471–1474 (2009)
27. Maestro (Schrödinger, LLC, New York, 2015)
28. Schrödinger (LLC, New York, NY 2016)
29. Desmond (D. E. Shaw Research, New York, 2015)
30. G. Chang, W.C. Guida, W.C. Still, J. Am. Chem. Soc. **111**, 4379–4386 (1989)
31. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, in Gaussian 09, Revision B.01 (Gaussian Inc., Wallingford, CT, 2010)

Chapter 9
Overall Perspective and Conclusions

Many techniques were utilized throughout this project, all focused on increasing the structural knowledge of organic and natural compounds, utilizing different parts of the NMR experimental toolbox. Large parts of the work have been focused on model compounds, usually strychnine. The true test to whether the reported methods, being NMR experiments (S^3 HMBCs) or RDC calculations or alignment medias (θ-method and chiral media), will lead to actual useful information still lies ahead; when data of novel and/or biologically or otherwise interesting structures are investigated. The compounds of actual interest, natural products and synthetic peptides, were all investigated using NOEs and $^3J_{HH}$-coupling constants. These observables enabled the determination of stereochemistry of novel compounds that would have been difficult by use of a purely qualitative analysis, and led to novel insights into molecular structures in solution.

The first part of Chap. 3 concerns the elucidation of 3D structural features, primarily stereochemistry, of the novel natural compounds from fungal sources. While some of the structures might have been solved from qualitative data, quantitative distances added to credibility and made flexible systems solvable. Flexible used in the most rigid definition, that is. Some order is needed to gain a handle on the compounds, but the multi-structure approach, coupled to the sturdiness of the NOEs and orthogonality of the *J*-coupling constants, led to interesting results. A slight overfitting is inevitable, but the alternative of using too few structures is undesirable and limits information to a level where one could utilize only computational chemistry—and modern spectrometers and experiments are sturdy enough to avoid major errors. The natural products are followed by synthetic peptides of biological importance. Again NOEs and *J*-coupling constants led to structural knowledge, as either structural differences between the three HDAC inhibiting azumamides, including more knowledge of the conformational space in solution, or knowledge of the interactions between two peptides that exhibit a high degree of specific structural recognition. In the structural recognition, back-calculated distances were not the prime source of information. Instead the distances were used to limit conformational space in simulations—leading to two

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9_9

complimentary structures. An important lesson learned was, that while very rigid and ordered structures may be "boring" and "too easily solved", very flexible and chaotic structures are hardly solvable at all. In both cases NMR spectroscopy may help, but the approach has to be varied to fit the task at hand. The last part of the chapter focuses mostly on the biggest assumption taken, namely that the rotational correlation time is equal among all nuclei pairs. While the approaches utilized in the literature for molecules in aqueous solution may work well in giving an estimate to the rotational correlation time in organic solvents, care should be taken when trying to implement the rotational correlation time in calculations for organic solvent. This is largely still a work in progress.

The next chapter, Chap. 4, presents two novel NMR experimental approaches to obtain homo- or heteronuclear long-range coupling constants. These S^3 HMBC experiments led to easily interpretable spectra and very precise extraction of coupling constants. Especially the homonuclear variant is unrivaled in the information obtained, while the larger amount of experiments for long-range heteronuclear coupling constants leads to the experiment being an excellent addition to an already extensive pool. The fact that only coupling constants to methine carbons may be extracted is limiting, but great sensitivity, ease of use and high correlation of the extracted coupling constants to theoretical values are significant upsides to the experiments. Both experiments are prime candidates for being used extensively in the field of carbohydrate chemistry, as methines are abundant and long-range coupling constants an already established important structural parameter, as well as more general structural elucidation.

The last three Chaps. (6–8) concern residual dipolar coupling constants (RDC) and their usage in structural elucidation of small molecules. The chapters may largely be seen as trying to solve the absolute and relative stereochemistry—the first for rigid structures and the latter for "flexible". Flexible is probably an overstatement, but the structures investigated were very flexible compared to the rigid compounds usually published. A considerable amount of long-range RDCs were extracted by utilizing S^3 HMBC *homo*, which may lead to a just as large amount of information, especially envisioned for more flexible small molecules. Enantiodiscrimination was achieved, but to a minor degree, using stretched polymers. While the prospect of utilizing stretched polymers for absolute stereochemical assignments is exciting, the current polymers do not perform to a degree where they will overtake the LC lead in the field anytime soon. But it is a start. The novel θ-method led to more information on conformer populations than the routinely utilized SVD for flexible compounds, or equal amount of information in much less time for less flexible structures. While this technique is not suited for certain tasks, e.g. enantiodiscrimination, a method of gaining population information from RDCs has been searched for, since they were introduced to small molecules. While better methods may emerge, the θ-method is easily implemented, relies on only the most simple of RDC formulae, and resulted in structural population information almost rivaling the NOEs. Not quite though, due to optimal averaging conditions for determining conformers of NOE distances.

All in all many challenges were faced and tried resolved. Some projects need more work, some are finished with structural knowledge gained, while other will need to stand the test of letting other users try to utilize the methods, hopefully with success. The S^3 HMBC experiments are "plug and play" experiments, which will hopefully lead to an immediate utilization in the small molecular NMR community. The θ-method is, though promising, not expected to replace SVD and other tensor based methods due to assumptions taken up front, but may be used as a complementary method in NOE and *J*-coupling constants investigations—easily interfaced in a coupled approach to gain structural insights.

Appendix

The Appendices was shortened in the final print of the thesis, and only Appendix A1, A2, A3, A5 and A10 are included. The remaining Appendices were submitted to the Ph.D. Committee, and the full Appendix is available upon request to the e-mail addresses: chg@kemi.dtu.dk

A.1 NMR

A.1.1 Spectrometers

DTU

All spectra were acquired at one of the five NMR spectrometers given below. Spectra were acquired using standard pulse sequences at 25 °C unless stated otherwise. The spectra were recorded using 5 mm NMR tubes when possible (solvent volume 500 μL) or 3 mm tubes (solvent volume 180 μL). Chemical shifts are always given in ppm, and coupling constants in Hz.

- Varian Inova 500 MHz with a 5 mm HCP probe (499.87 MHz for ^{1}H, 125.70 MHz for ^{13}C).
- Bruker Avance III 400 MHz with a 5 mm BBO smartprobe (400.23 MHz for ^{1}H, 100.64 MHz for ^{13}C).
- Bruker Ascend 400 MHz with a 5 mm H-br.band dual channel z-gradient Prodigy cryoprobe (400.13 MHz for ^{1}H, 100.61 MHz for ^{13}C).
- Bruker DRX 600 MHz with a 5 mm BBO smartprobe (600.13 MHz for ^{1}H, 150.92 MHz for ^{13}C).
- Bruker Avance 800 MHz with a 5 mm TCI cryoprobe (798.80 MHz for ^{1}H, 200.86 MHz for ^{13}C).

C. R. Hoeck, *Solving a 3D Structural Puzzle*, Springer Theses,
https://doi.org/10.1007/978-3-319-96172-9

External stay

The NMR spectrometers used for acquisitions at the University of Bristol are given below.

- Varian VNMRS DirectDrive 500 MHz with a broadband two-channel OneNMR probe (499.66 MHz for ^{1}H, 125.64 MHz for ^{13}C)
- Varian VNMRS DirectDrive 600 MHz spectrometer equipped with an indirect observe cryoprobe (600.05 MHz for ^{1}H, 150.88 MHz for ^{13}C).

A.1.2 Solvents

The solvents used are given below along with the solvent resonances used for referencing the spectra [1]. Chemical shifts (δ) are in ppm and scalar couplings are reported in hertz (Hz).

Solvent	δ_H (ppm)	δ_C (ppm)
DMSO-d_6	2.50	39.5
$CDCl_3$	7.26	77.2
CD_3OD	3.31	49.0
D_2O	4.79	–

A.1.3 Acquisition

Examples of the basic setup of the most used literature experiments are found below. The pulse sequence given are examples of sequences often used on Bruker spectrometers.

1D ^{1}H experiments (w/wo water suppression) were used to acquire 1D proton spectra. A standard acquisition was as follows: si (Fourier transform size) = 64k, rd (relaxation delay) = 1.0 s, ns (number of scans) = 32, ds (number of dummy scans) = 8. (**zg30**)

CLIP-HSQC experiment was used to acquire f2-coupled HSQC spectra for the extraction of $^{1}J_{CH}$ and $^{1}T_{CH}$ (RDCs). A standard acquisition was as follows: si = 4096, ni (number of increments) = 256, rd = 1.0 s, ns = 8, ds = 16. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension. Coupling constants were extracted by overlay of 1D slices through F1. (**CLIP_hsqcetgp**, pulse sequence acquired from http://www.ioc.kit.edu/luy/110.php website of Prof. Dr. Burkhard Luy, Institut für Organische Chemie, Karlsruher Institut für Technologie) [2].

NOESY experiments (w/wo water suppression) were used to acquire 2D NOESY spectra. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8, ds = 32. The mixing time was usually 50–300 ms and the relaxation delay depends on the T_1 of the protons (usually around 1–5 s used). The spectra were

zero filled to 8k in the F2 and 1k in the F1 dimension. Integrals were determined from the spectra in Topspin. (**noesyp, noesyphpr, noesygpph19**) [3–6].

ROESY experiments (w/wo water suppression) were used to acquire 2D T-ROESY spectra. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8, ds = 32. The mixing time was usually 50–300 ms and the relaxation delay depends on the T_1 of the protons (usually around 1–5 s used). The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. Integrals were determined from the spectra in Topspin. (**roesyph.2, roesyphpr.2, roesygpph19.2**) [5–8].

DPFGSE NOESY (double-pulsed-field-gradient-spin-echo) or **SPFGSE NOESY** (single-pulsed-field-gradient-spin-echo) experiment was used for 1D NOESY experiments. A standard acquisition was as follows: si = 64k, rd = 1.0 s, ns = 128, ds = 8. The mixing time was usually 300–500 ms. The spectra were zero filled to 128k. Integrals were determined by setting the integral of the irradiated resonance equal to −1000 and determining the other relative integrals (PANIC approach).

DQF-COSY experiments were used to acquire 2D COSY spectra. A standard acquisition was as follows: si = 4096 or 8096, ni = 512, ns = 8, ds = 32, rd = 1.0 s. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**cosygpmfphpp**) [9−12].

gHSQC experiments were used to acquire multiplicity edited HSQC spectra. A standard acquisition was as follows: si = 2048, ni = 256, ns = 8, ds = 32, rd = 1.0 s. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**hsqcedetgpsisp2.3**) [13−18].

gHMBC experiments were used to acquire HSQC spectra. A standard acquisition was as follows: si = 2048, ni = 256, ns = 8, ds = 32, rd = 1.0 s, J_{HMBC} = 8 Hz. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**hmbcetgpl3nd**) [19, 20].

S^3 HMBC homo experiments were used to acquire S^3 edited HMBC spectra for extraction of homonuclear coupling constants. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8 or 16, ds = 32, rd = 1.0 s, J_{HMBC} = 8 Hz. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension.

S^3 HMBC hetero experiments were used to acquire S^3 edited HMBC spectra for extraction of heteronuclear coupling constants. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8 or 16, ds = 32, rd = 1.0 s, J_{HMBC} = 8 Hz. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension.

The new pulse sequences of **S^3 HMBC homo/hetero** are not included, but are provided upon request, and the acquisition data are given in the thesis.

A.2 Other Equipment

Analytical RP-UPLC-MS analyses were performed on a Waer Aquity RP-UPLC system with a diode array detector (DAD) coupled to a SQD mass spectrometer. The ionization method was ESI. The column was a Aquity UPLC BEH C18 column

(1.7 μm, 2.1 × 50 mm, 0.6 mL/min, 65 °C) Eluents were water and MeCN (both added 0.1% HCO_2H), and a linear gradient of 5–100% MeCN over aprox. 2.6 min was used.

Analytical RP-UPLC-HRMS data were recorded on a Maxis 3G UHR-QTOF-MS by Bruker Daltonics with an UltiMate 3000 Rapid Separation LC by system Thermo Scientific Dionex. The ionization method was ESI and UV spectra from the DAD were collected at wavelengths from 200 to 700 nm. The column used was a Kinetex C18 (2.6 μm, 100 × 2.1 mm, 0.4 mL/min, 40 °C). Solvents were MeCN and water (both added 20 mM HCO_2H), and a linear gradient of 10–100% MeCN in 10 min followed by 100% MeCN for 3 min was used.

Optical rotation was measured on a Perkin-Elmer 341 polarimeter, with a 100 mm cell of 1 mL, with a sodium lamp (589 nm, 20 °C).

A.3 Software

A.3.1 Data Analysis

To analyse obtained HPLC-DAD-MS spectra, the program Data Analysis by Bruker was used. An internal standard was used to gain higher mass accuracy. Smart Formula was used to generate and evaluate constituent formulas for a given mass.

A.3.2 Gaussian

The program Gaussian version 09 revision B.01 by Gaussian was used for DFT calculations including optimizations and NMR calculations [21].

A.3.3 Maestro

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations [22]. The programs MacroModel version 10.8 and Desmond version 4.2 were used [23, 24].

A.3.4 Matlab

The program Matlab version R2015a (2015) by The MathWorks Inc. was used for scripting and automating processes. The scripts were all made in house utilizing the general available scripts imbedded in the program.

The script RBNMR by Niels Nyberg was used to load NMR spectral data into Matlab.

A.3.5 MestReNova

The obtained 1D and 2D NMR spectra from my external stay were processed in MestReNova v. 6.2.1 by MestReLab Research S. L.

A.3.6 MSPIN

The software MSPIN version 1.3.3-79 (2013) by MestReLab Research S. L. was generally used for RDC (SVD) back-calculations [25]. The program may also be used for NOE and *J* coupling constant calculation but this was not generally used, though it was used to check the HLA calculations of the scripts. The NOE calculations were not used as the program uses a matrix approach.

A.3.7 PALES

The RDC software PALES was initially used and compared to MSPIN [26]. The latter program was used throughout and the two programs were found to give comparable results.

A.3.8 Topspin

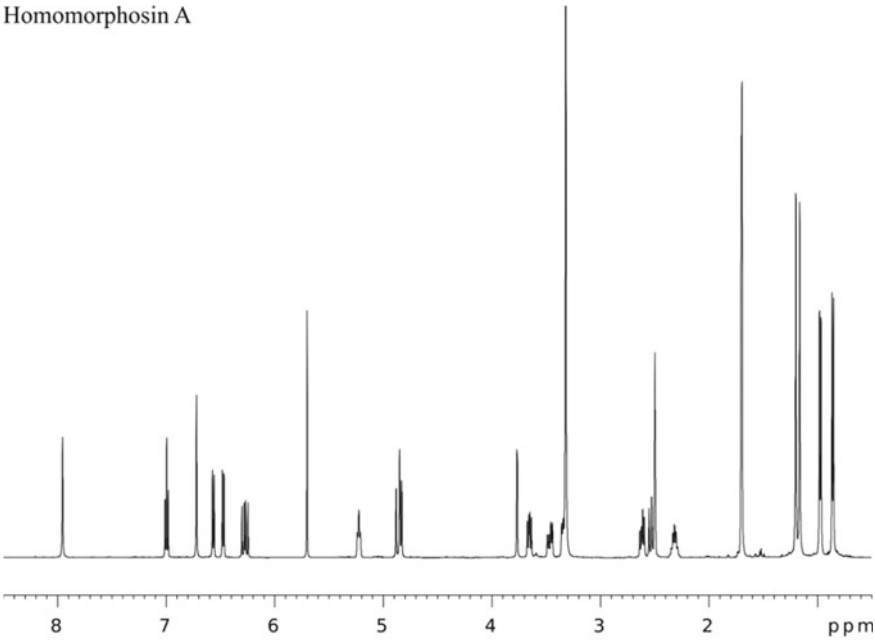
The obtained 1D and 2D NMR spectra were generally processed in Topspin version 3.1 (2012) by Bruker BioSpin.

A.5 NMR Spectra

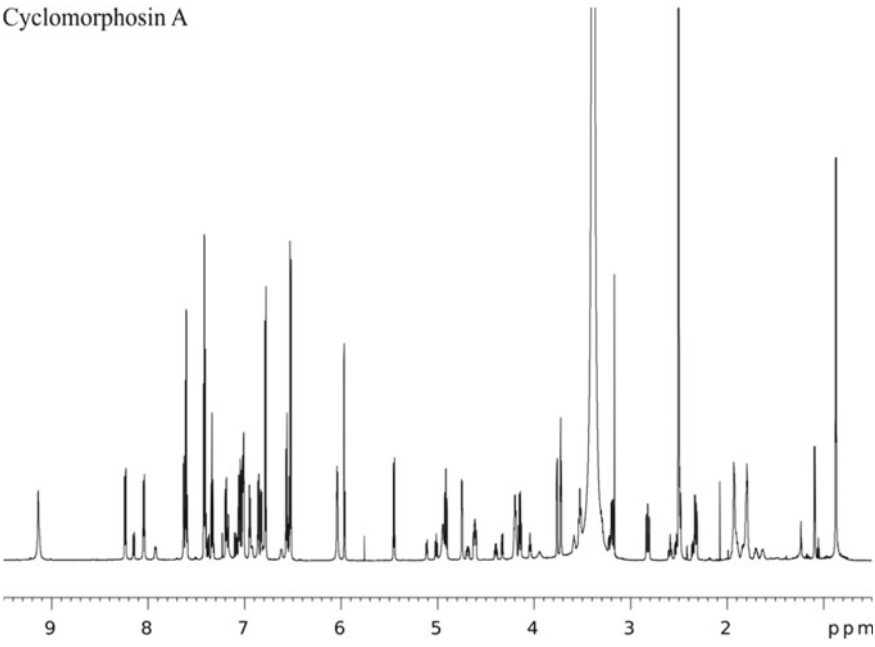
A.5.1 1D Spectra

A.5.1.1 Natural Compounds

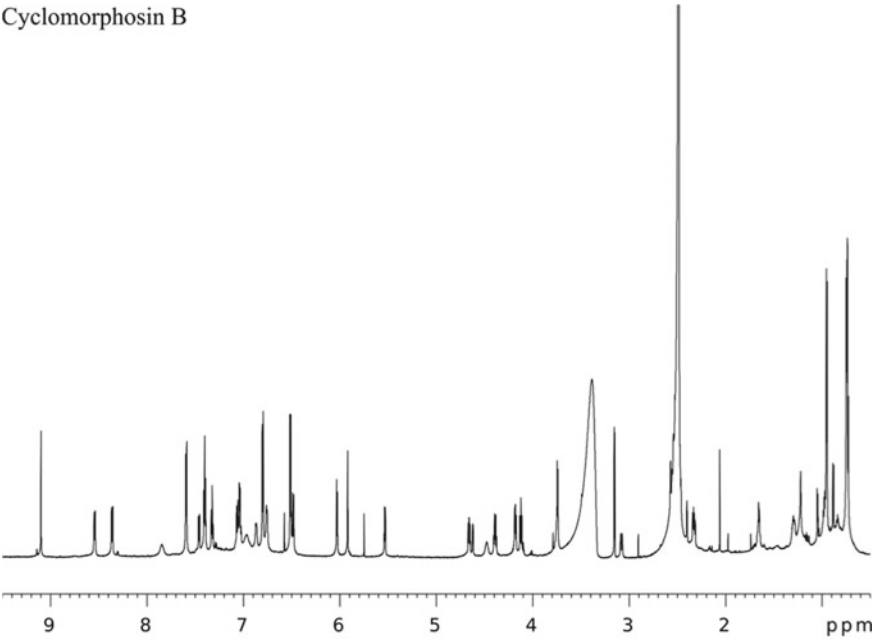
Homomorphosin A



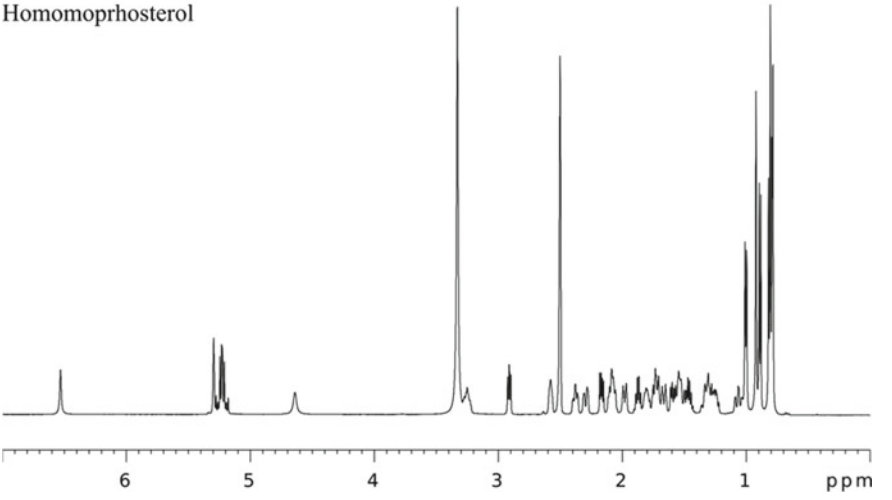
Cyclomorphosin A



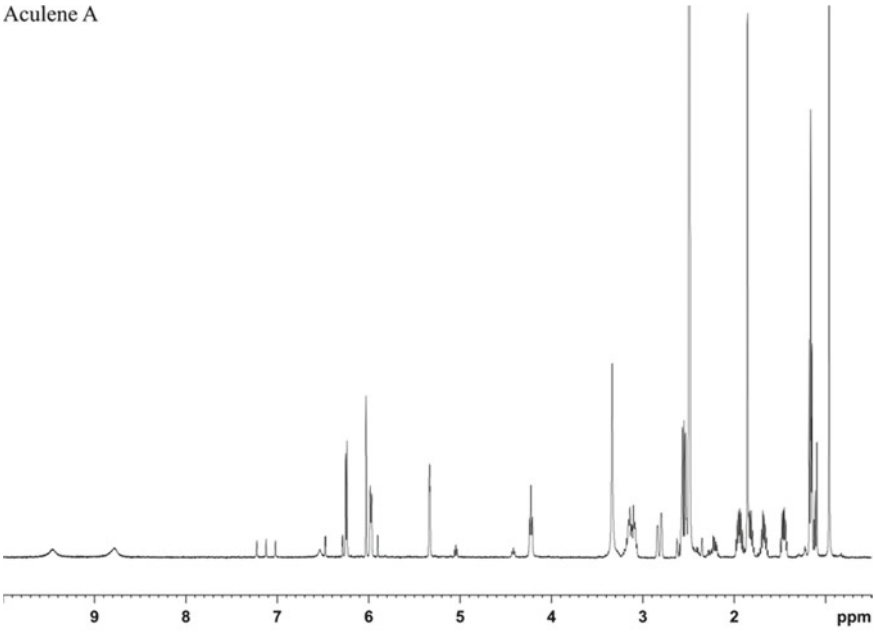
Cyclomorphosin B



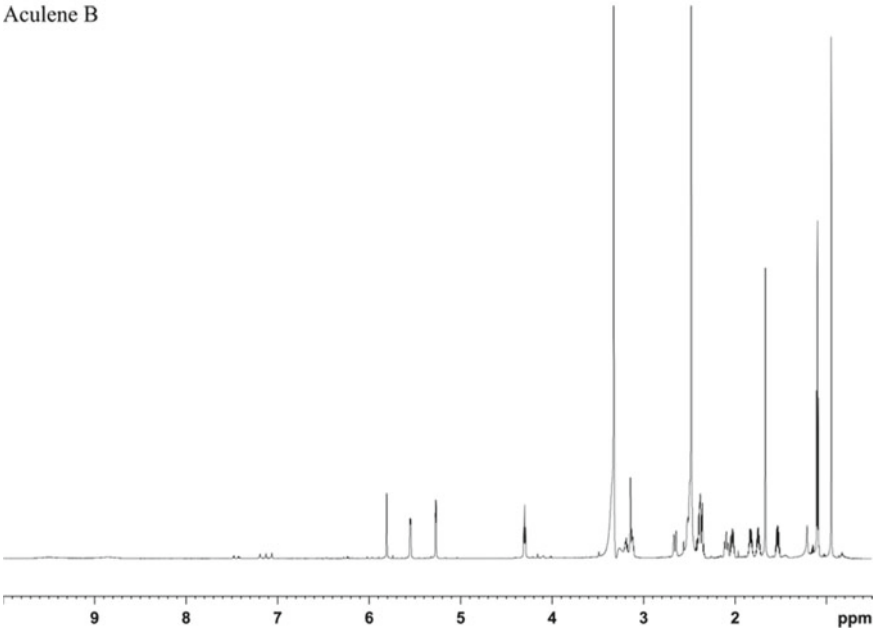
Homomoprhosterol



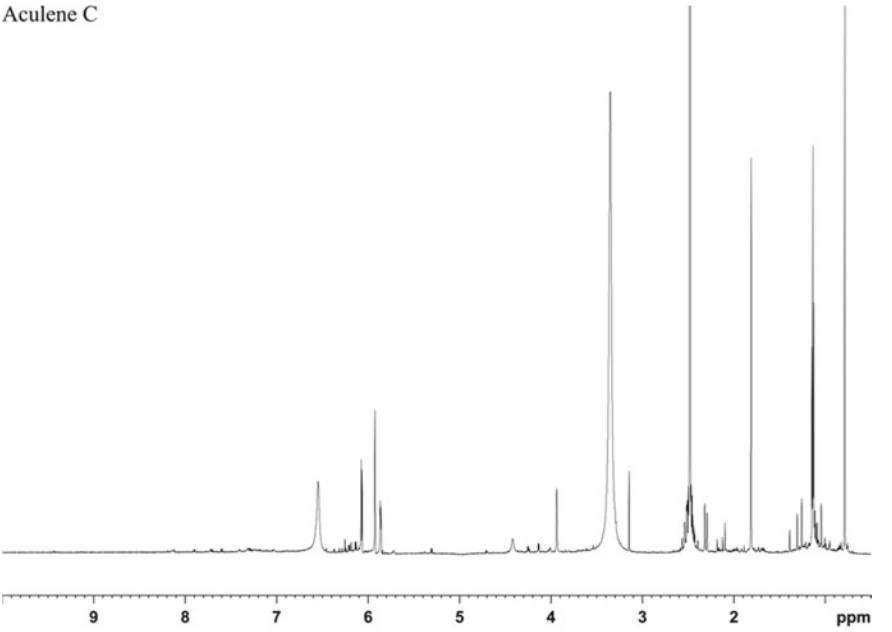
Aculene A



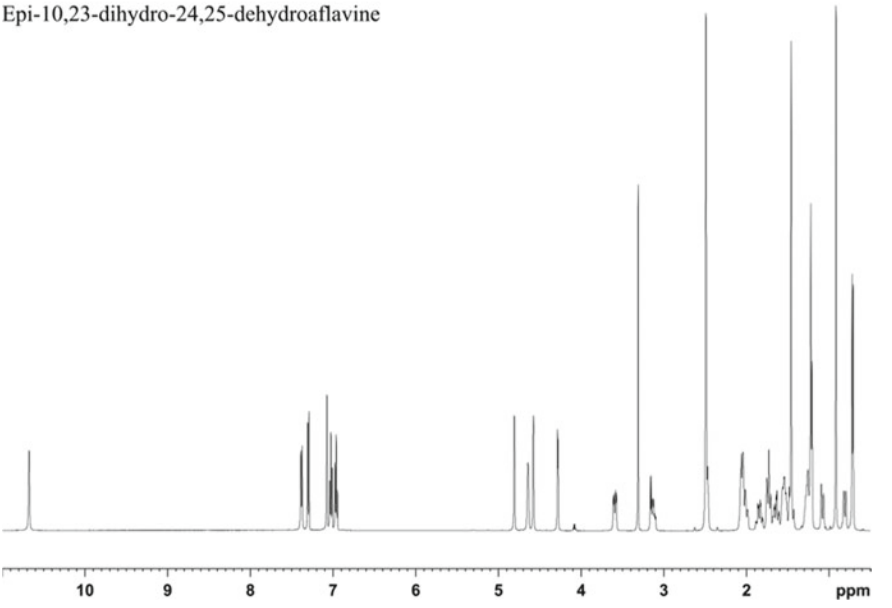
Aculene B



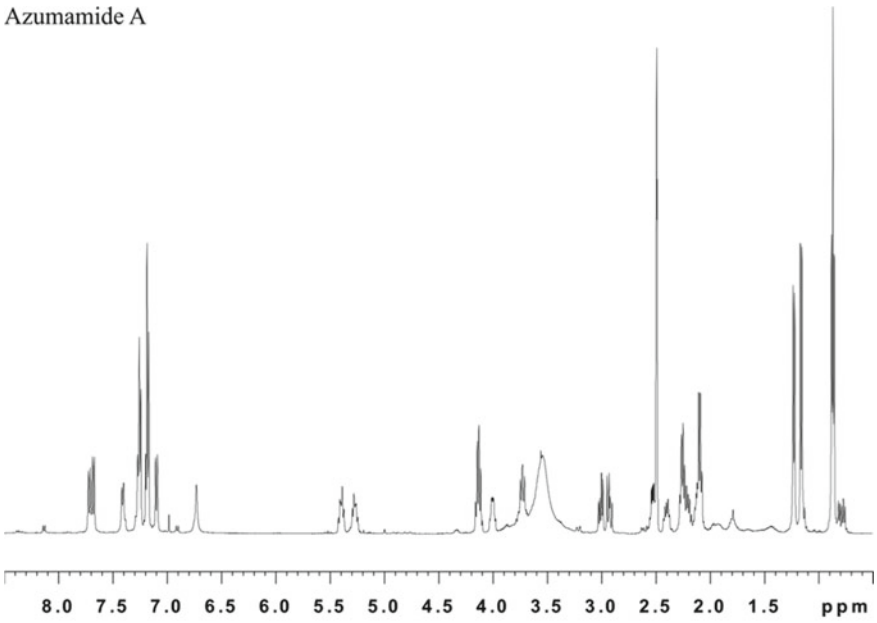
Aculene C

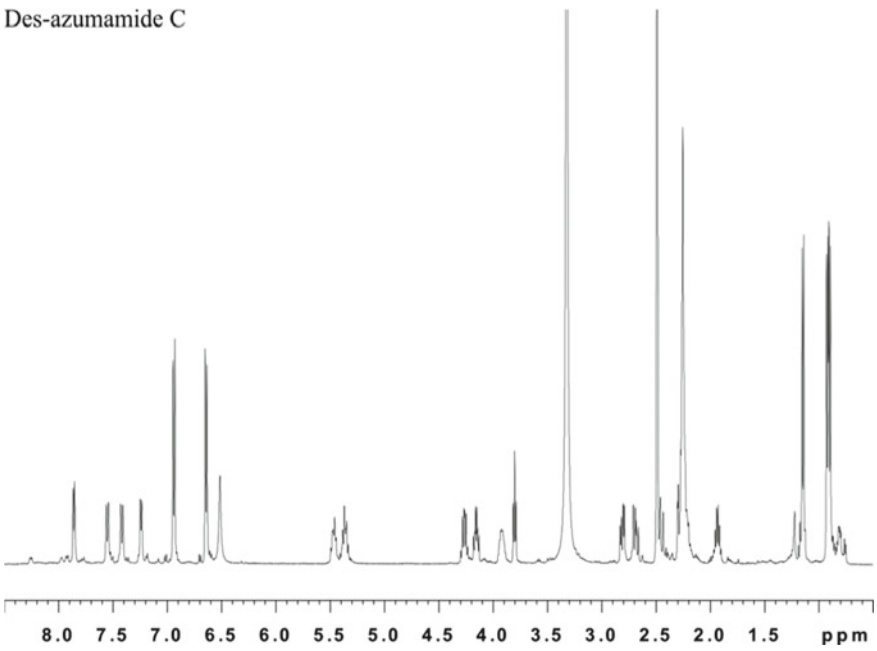


Epi-10,23-dihydro-24,25-dehydroaflavine

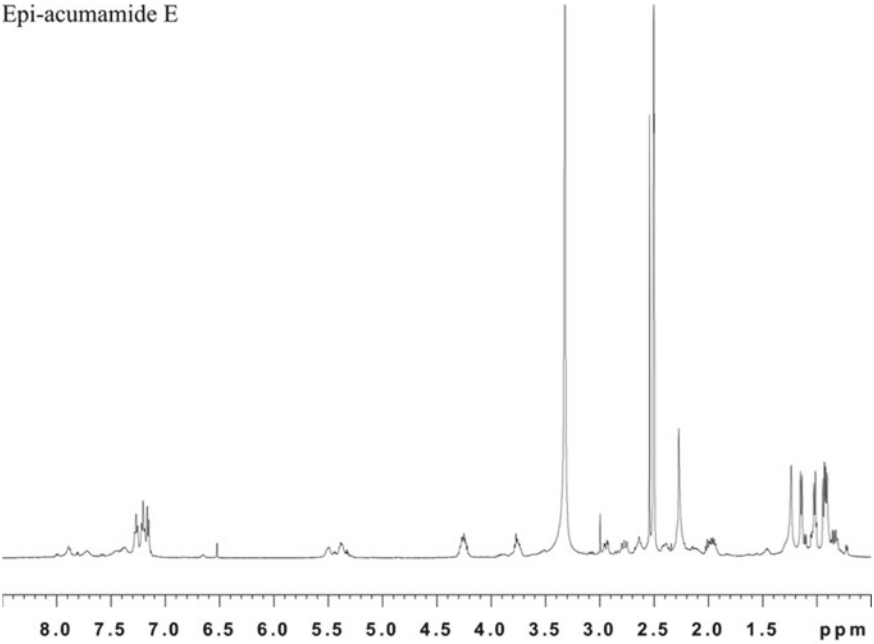


A.5.1.2 Synthetic Compounds

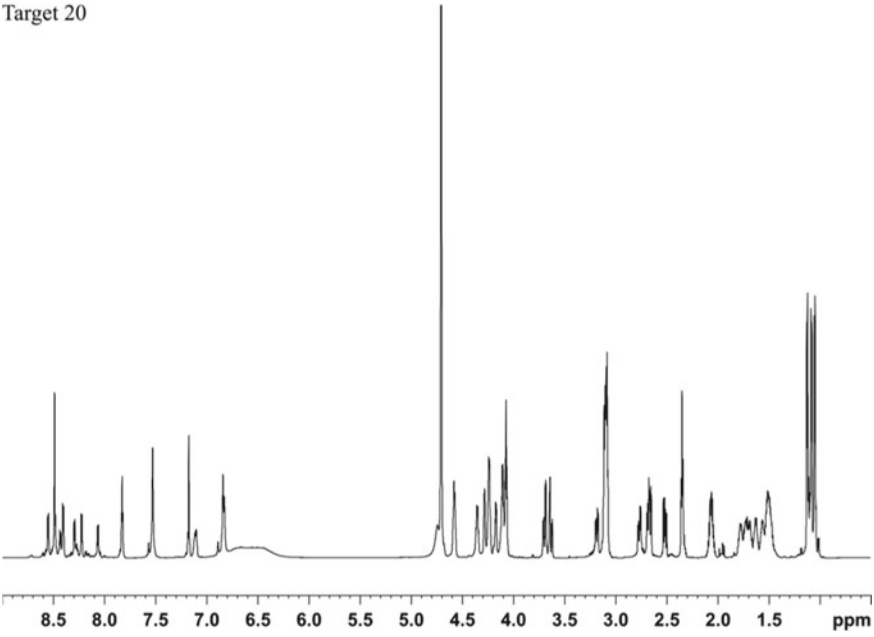




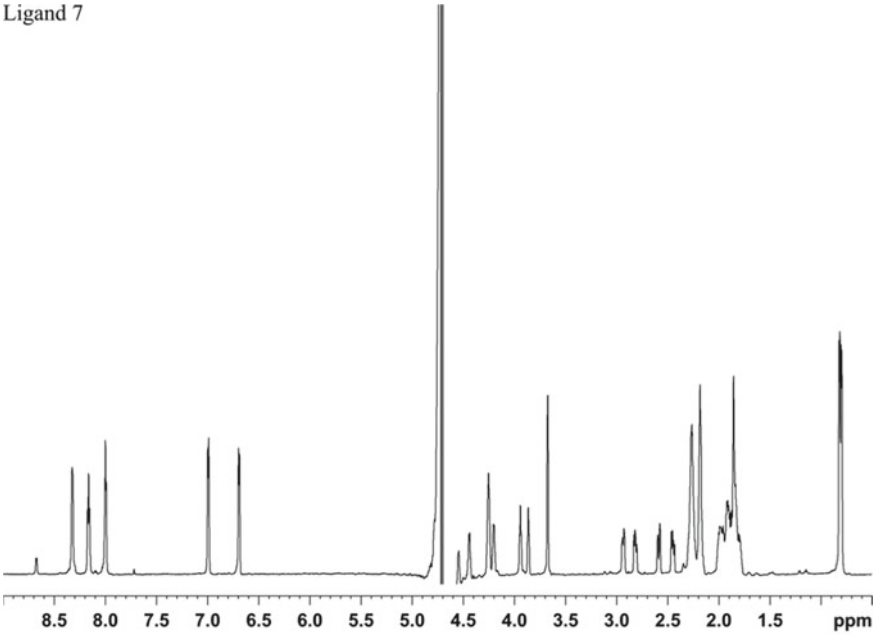
Epi-acumamide E



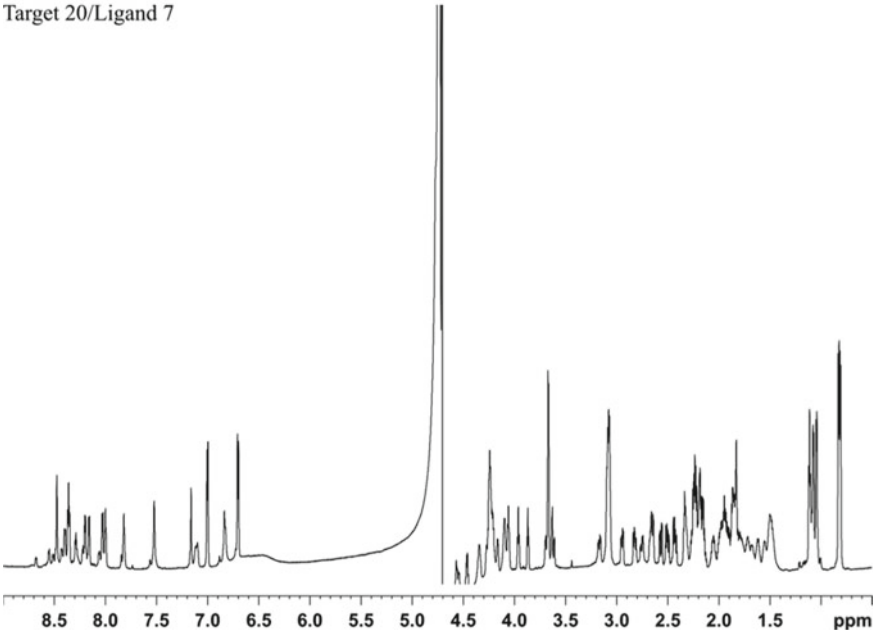
Target 20



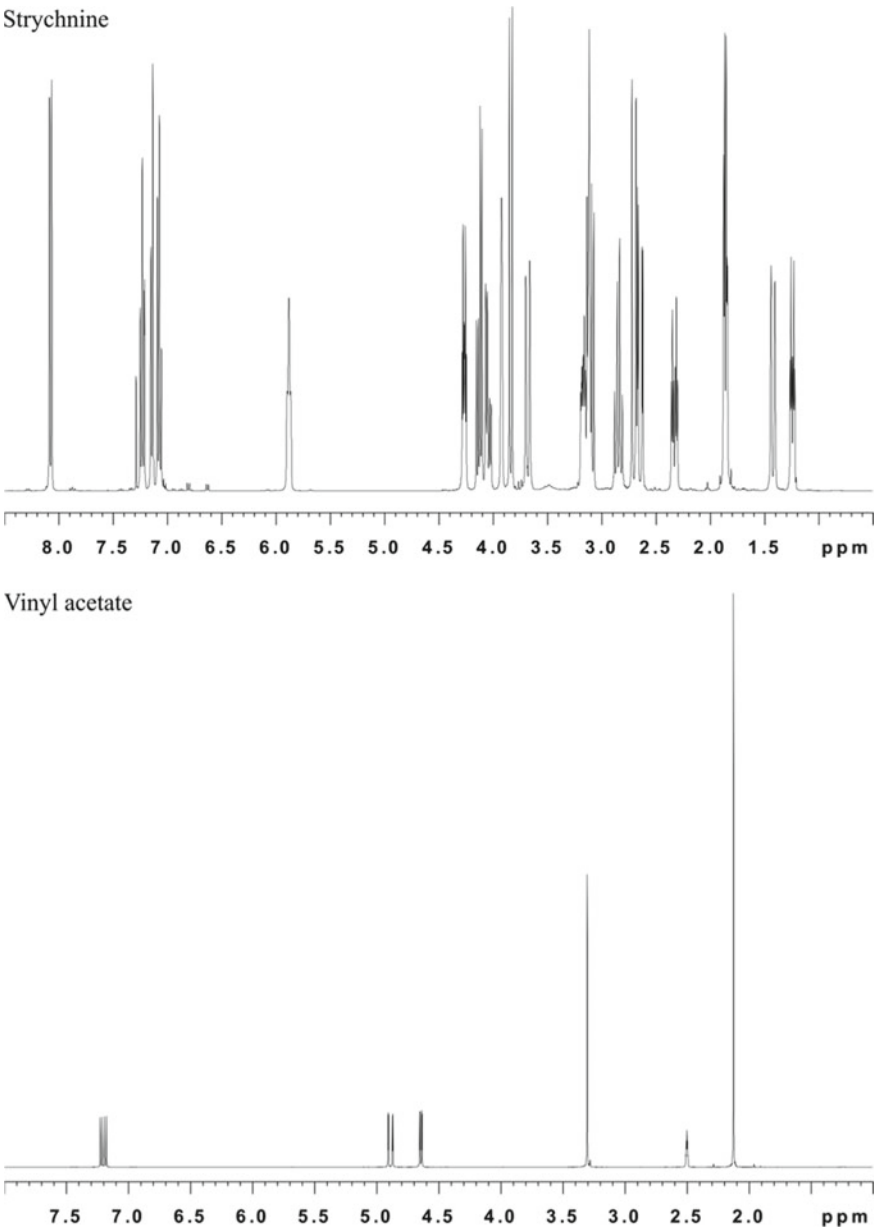
Ligand 7



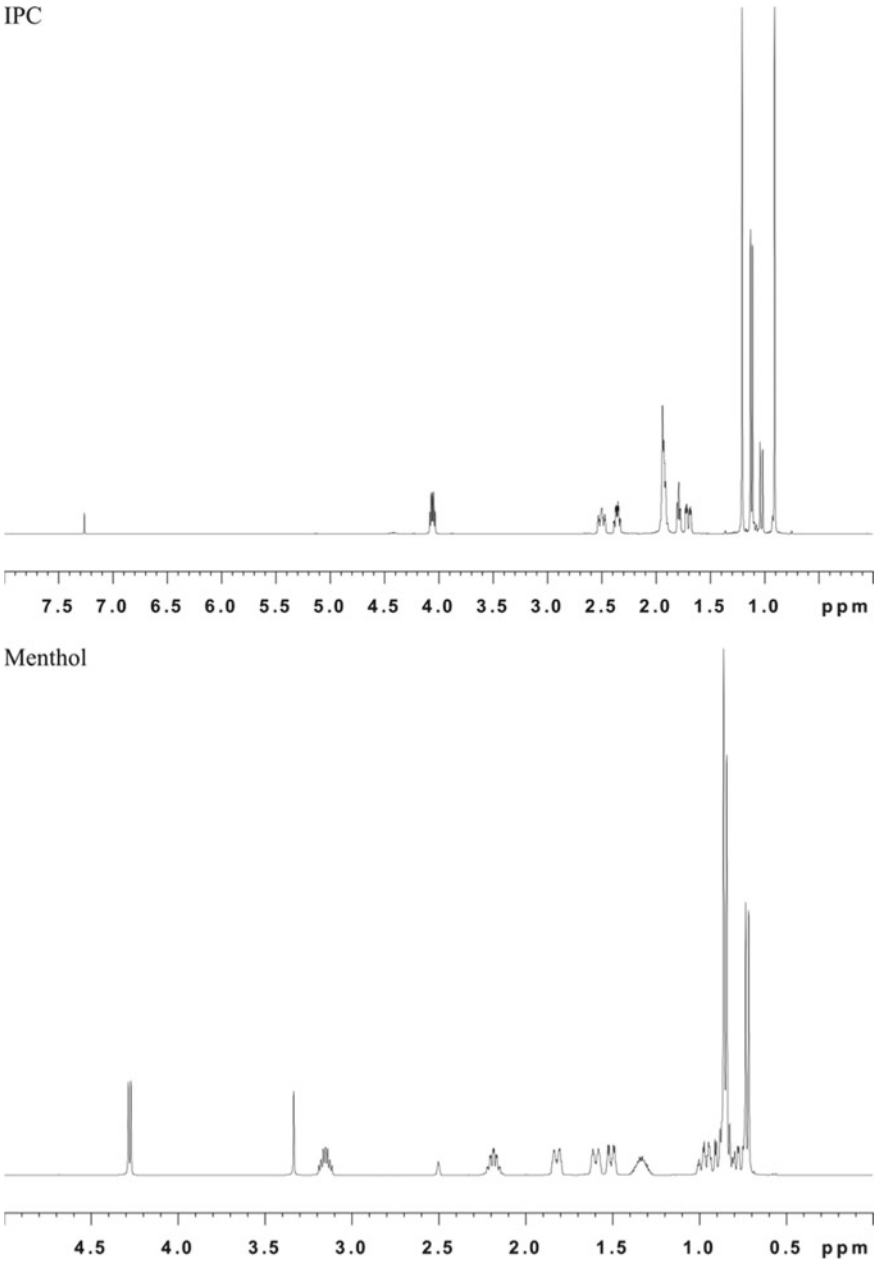
Target 20/Ligand 7



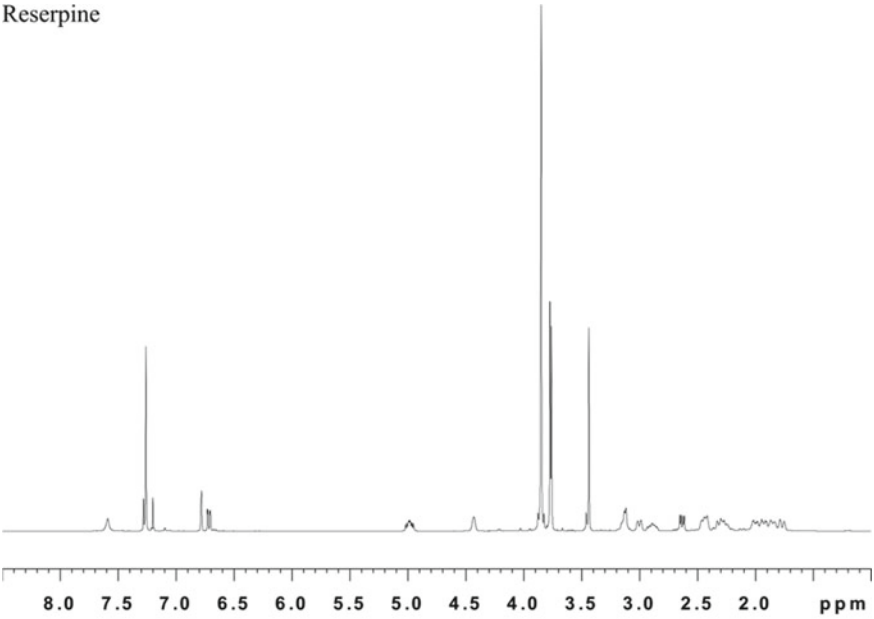
A.5.1.3 S^3 HMBC


Strychnine
8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 ppm
Vinyl acetate
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 ppm

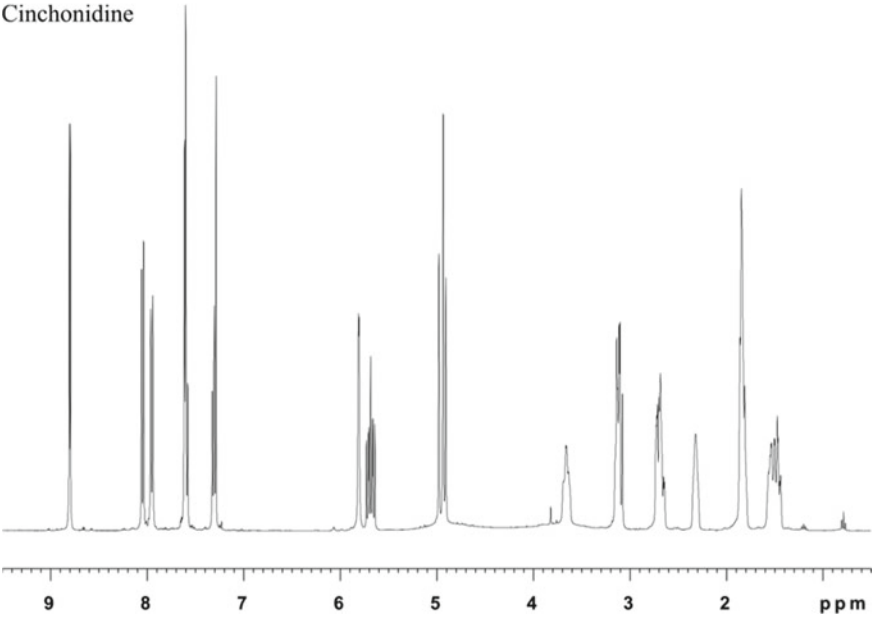
A.5.1.4 RDC


IPC
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm
Menthol
4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

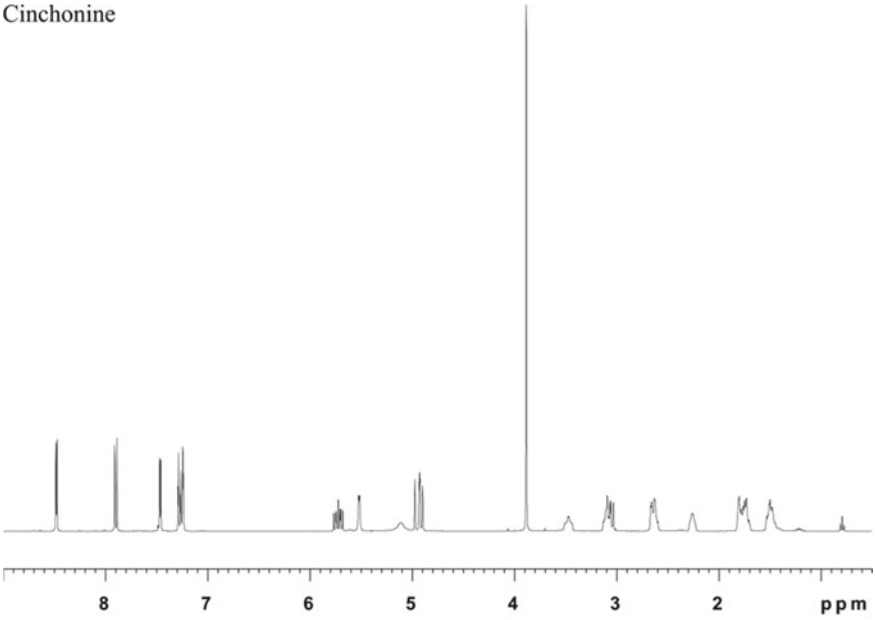
Reserpine



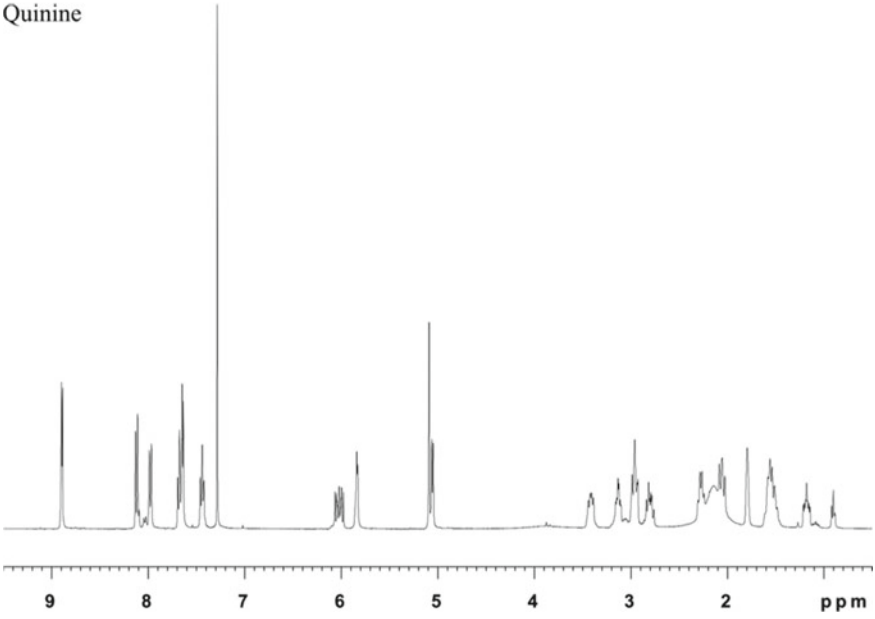
Cinchonidine



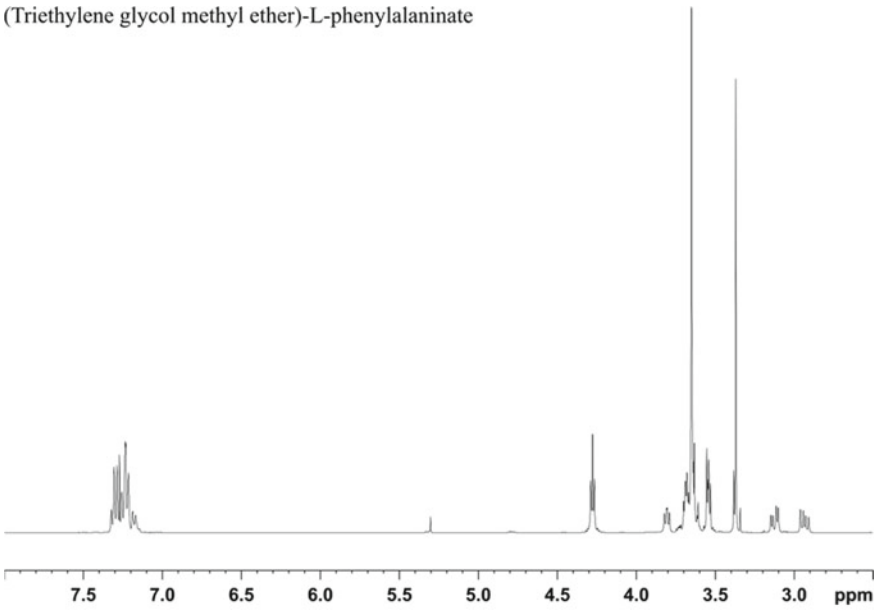
Cinchonine

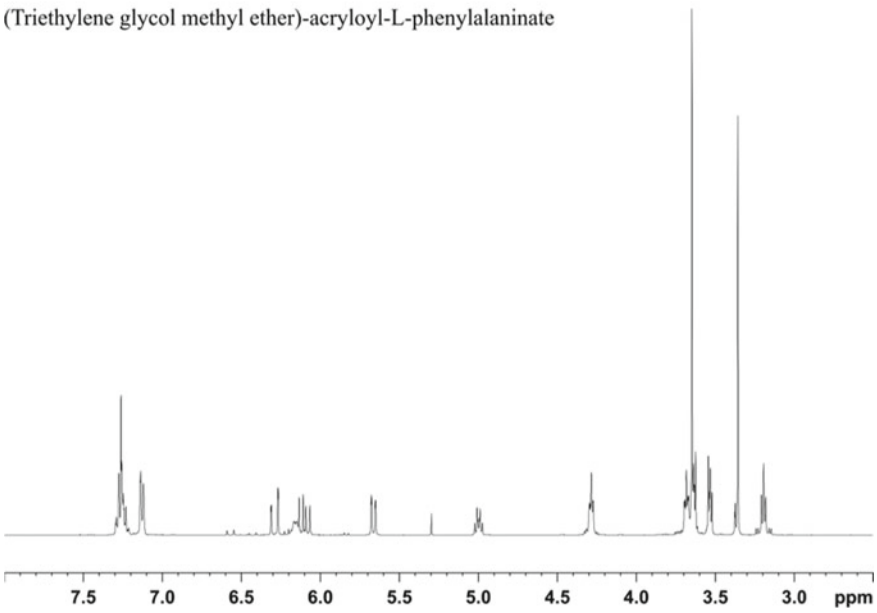


Quinine

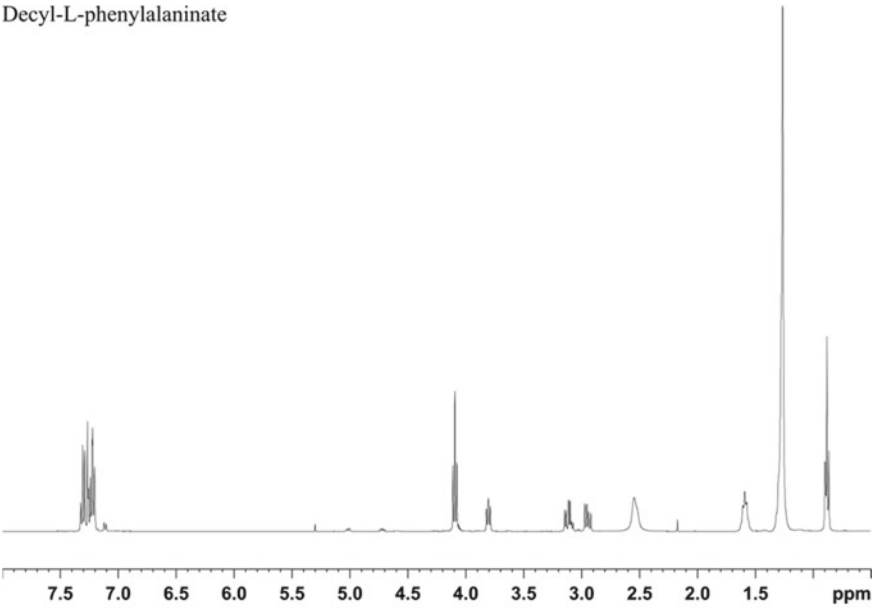


A.5.1.5 Synthesized Monomers

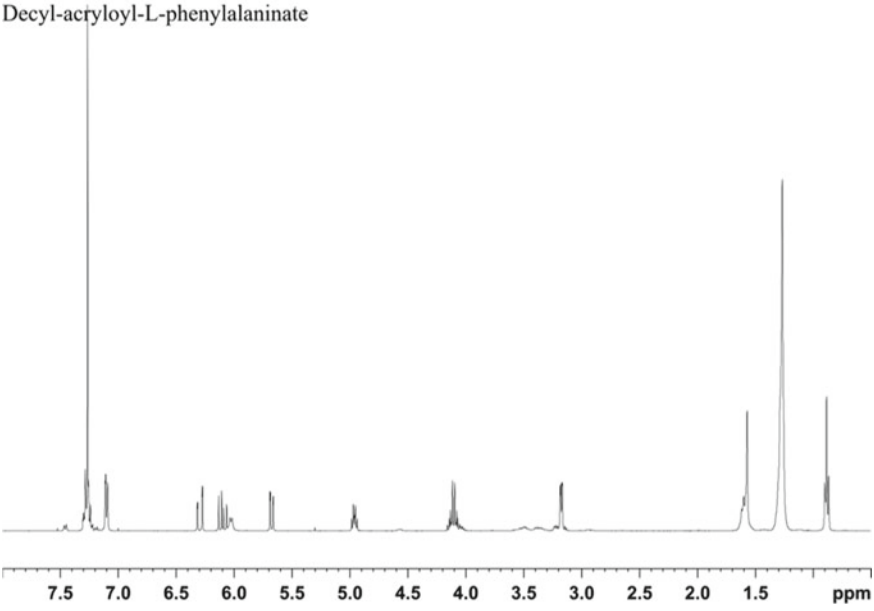




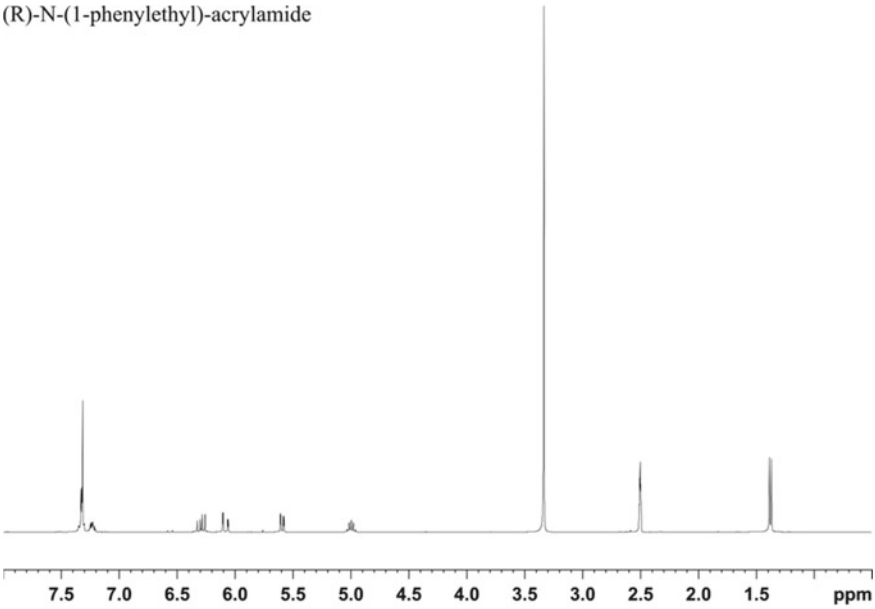
Decyl-L-phenylalaninate



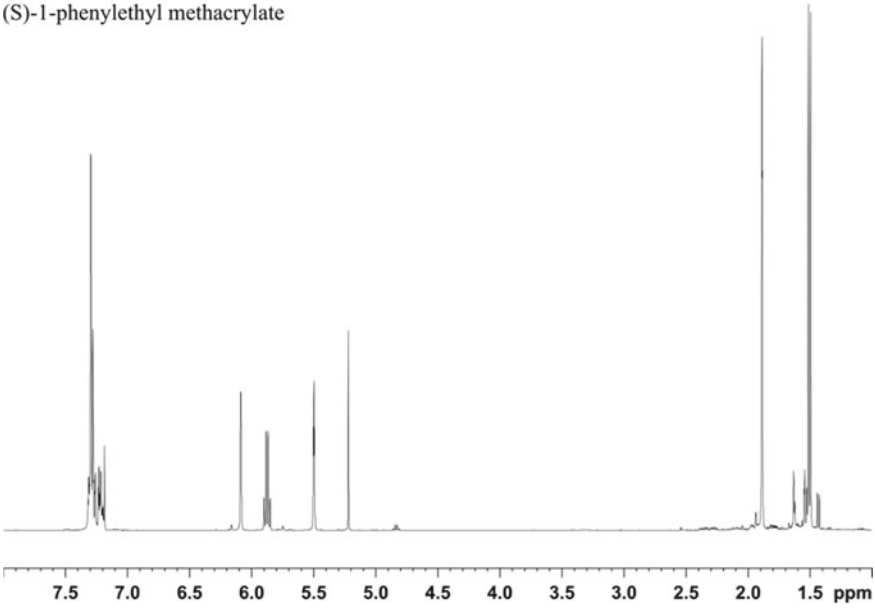
Decyl-acryloyl-L-phenylalaninate



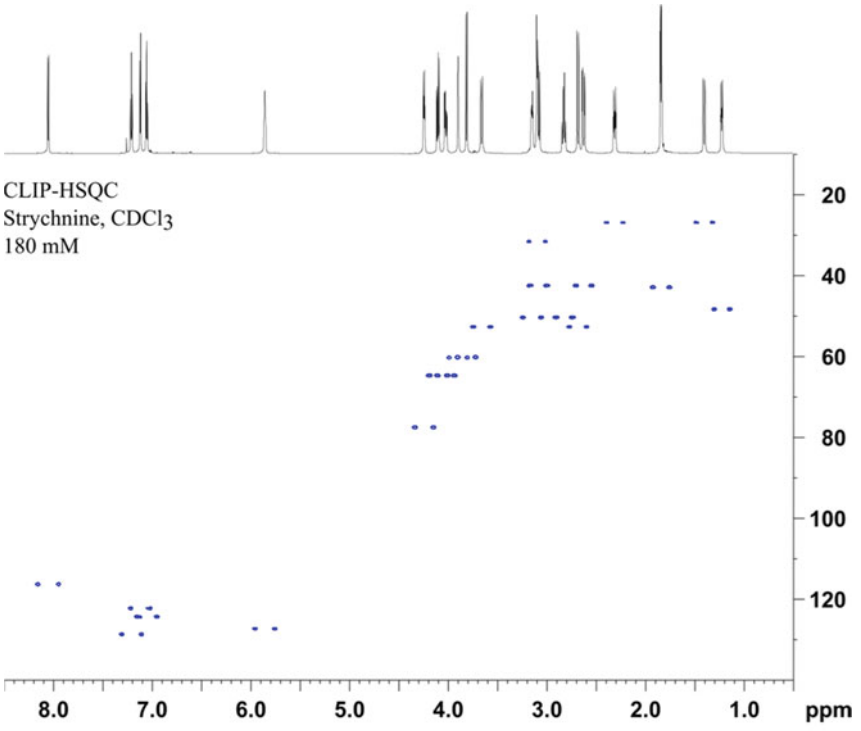
(R)-N-(1-phenylethyl)-acrylamide

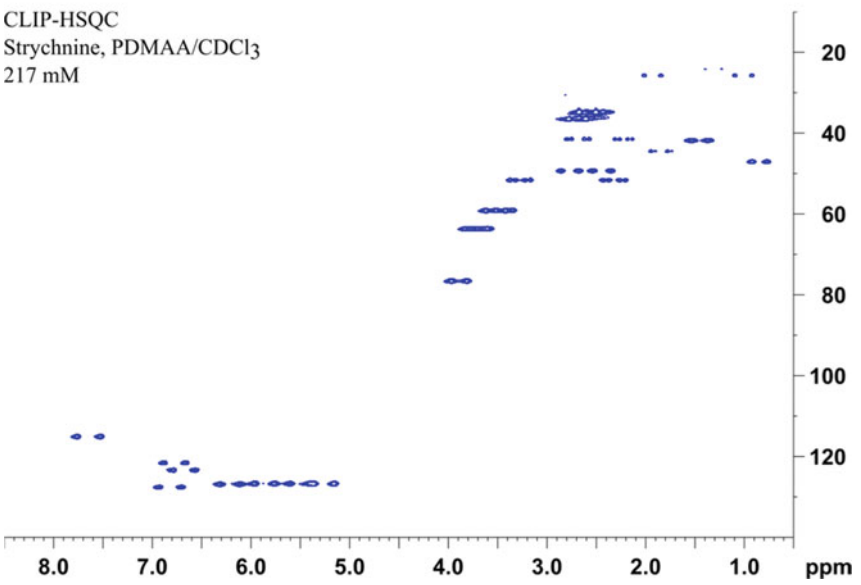


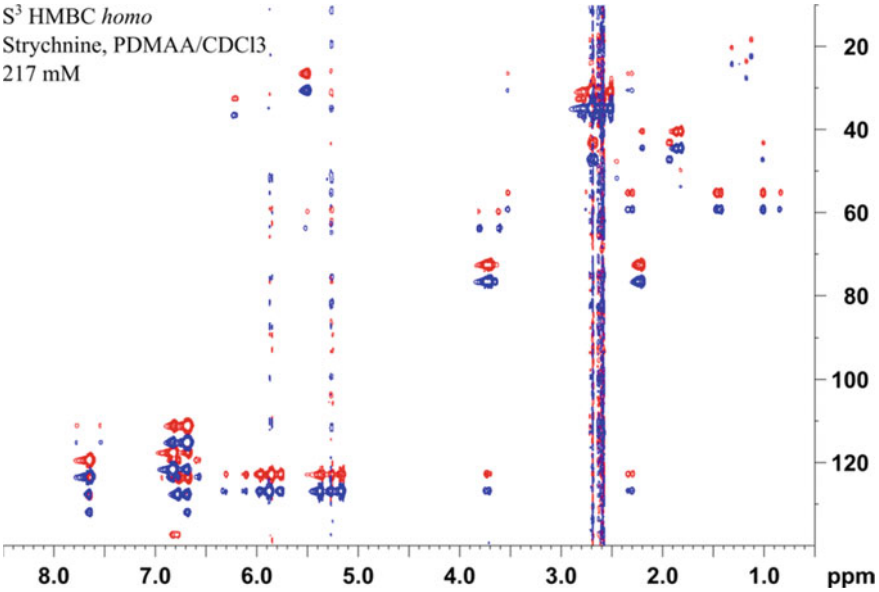
(S)-1-phenylethyl methacrylate



A.5.2 Examples of Isotropic and Aligned Spectra







A.10 References

1. H.E. Gottlieb, V. Kotlyar, A. Nudelman, J. Org. Chem. **62**(21), 7512–7515 (1997)
2. A. Enthart, J.C. Freudenberger, J. Furrer, H. Kessler, B. Luy, J. Magn. Reson. **192**, 314–322 (2008)
3. J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, J. Chem. Phys. **71**, 4546–4553 (1979)
4. A. Kumar, R.R. Ernst, K. Wüthrich, Biochem. Biophys. Res. Commun. **95**, 1–6 (1980)
5. M. Piotto, V. Saudek, V. Sklenar, J. Biomol. NMR **2**, 661–666 (1992)
6. V. Sklenar, M. Piotto, R. Leppik, V. Saudek, J. Magn. Reson. Ser. A **102**, 241–245 (1993)
7. T.L. Hwang, A.J. Shaka, J. Am. Chem. Soc. **114**, 3157–3159 (1992)
8. A. Bax, D.G. Davis, J. Magn. Reson. **63**, 207–213 (1985)
9. W.P. Aue, E. Bartholdi, R.R. Ernst, J. Chem. Phys. **64**, 2229–2246 (1976)
10. U. Piantini, O.W. Sørensen, R.R. Ernst, J. Am. Chem. Soc **104**, 6800–6801 (1982)
11. R.E. Hurd, J. Magn. Reson. **87**, 422–428 (1990)
12. A.L. Davis, E.D. Laue, J. Keeler, D. Moskau, J. Lohman, J. Magn. Reson. **94**, 637–644 (1991)
13. L.E. Kay, P. Keifer, T. Saarinen, J. Am. Chem. Soc. **114**, 10663–10665 (1992)
14. A.G. Palmer III, J. Cavanagh, P.E. Wright, M. Rance, J. Magn. Reson. **93**, 151–170 (1991)
15. W. Willker, D. Leibfritz, R. Kerssebaum, W. Bermel, Magn. Reson. Chem. **31**, 287–292 (1993)
16. J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S.J. Glaser, O.W. Sorensen, C. Griesinger, J. Biomol. NMR **4**, 301–306 (1994)
17. C. Zwahlen, P. Legault, S.J.F. Vincent, J. Greenblatt, R. Konrat, L.E. Kay, J. Am. Chem. Soc. **119**, 6711–6721 (1997)
18. R.D. Boyer, R. Johnson, K. Krishnamurthy, J. Magn. Reson. **165**, 253–259 (2003)
19. D.O. Cicero, G. Barbato, R. Bazzo, J. Magn. Reson. **148**, 209–213 (2001)
20. A. Bax, M.F. Summers, J. Am. Chem. Soc. **108**, 2093–2094 (1986)

21. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E.S, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B.M, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P.H, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M.H, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T.N, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, J, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox (Gaussian, Inc., Wallingford, CT, 2010)
22. Schrödinger, *Maestro* (LLC, New York, 2015)
23. Schrödinger, *Macromodel* (LLC, New York, 2015)
24. D.E. Shaw Research, *Desmond* (New York, 2015)
25. A. Navarro-Vázquez, Magn. Reson. Chem. **50**, S73–S79 (2012)
26. M. Zweckstetter, Nat. Protoc. **3**, 679–690 (2008)

Bibliography

27. J. Keeler, *Understanding NMR Spectroscopy*, 2nd edn. (Wiley, New Jersey, 2010)
28. O.W. Sørensen, Prog. Nucl. Magn. Reson. Spectrosc. **21**, 503–569 (1989)
29. J. Clayden, N. Greeves, S. Warren, P. Wothers, *Organic Chemistry* (Oxford University Press, Oxford, 2001)
30. P. Hubbard, W.J. Brittain, J. Org. Chem. **63**, 677–683 (1998)