

Werner Dubitzky · Jennifer Southgate  
Hendrik Fuß *Editors*

# Understanding the Dynamics of Biological Systems

Lessons Learned  
from Integrative Systems Biology

# Understanding the Dynamics of Biological Systems



Werner Dubitzky • Jennifer Southgate  
Hendrik Fuß  
Editors

# Understanding the Dynamics of Biological Systems

Lessons Learned from Integrative  
Systems Biology

*Editors*

Werner Dubitzky  
Nano Systems Biology Research Group  
Biomedical Sciences Research Institute  
University of Ulster  
Coleraine BT52 1SA  
United Kingdom  
[w.dubitzky@ulster.ac.uk](mailto:w.dubitzky@ulster.ac.uk)

Hendrik Fuß  
Nano Systems Biology Research Group  
Biomedical Sciences Research Institute  
University of Ulster  
Coleraine BT52 1SA  
United Kingdom  
[hendrik.fuss@gmail.com](mailto:hendrik.fuss@gmail.com)

Jennifer Southgate  
Department of Biology  
Jack Birch Unit of Molecular Carcinogenesis  
University of York  
York YO10 5DD  
United Kingdom  
[js35@york.ac.uk](mailto:js35@york.ac.uk)

ISBN 978-1-4419-7963-6 e-ISBN 978-1-4419-7964-3

DOI 10.1007/978-1-4419-7964-3

Springer New York Dordrecht Heidelberg London

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

*Systems biology* could be defined as the *quantitative* analysis of the *dynamic* interactions among several components of a biological system and aims to understand the behavior of the system as a *whole*. R&D in systems biology involves the development and application of systems theory concepts for the study of complex biological systems through iteration over *mathematical modeling* and *computational simulation* and *biological experimentation*. Systems biology could be viewed as a tool to increase understanding of biological systems and to develop more directed experiments and finally allow predictions.

The field of systems biology arose out of a biological problem which is essentially entailed by the complexity of biological life. It was created because of the limitations of conventional (reductionistic) biology in the investigation and understanding of complex biological phenomena arising from the dynamic interaction of many biological compounds. At present, a large number of individual genes or proteins which play key roles in essential physiological processes are known. For many of these, structural data and detailed mechanistic descriptions at a molecular level are available. In most cases, however, the individual characterization of these molecules is not sufficient to fully understand their immediate or their superordinate physiological function. Similarly, large networks of genes, proteins and other organic molecules have been discovered, mapped and characterized. While underlying mechanisms have been regarded as a promising base for explaining the multitude of cellular functions and phenomena observed *in vivo*, there is still a fundamental gap between the knowledge of a molecular mechanism and the understanding of the corresponding cellular or higher-level function.

The growing field of systems biology promises to bridge our current gap in understanding. Systems biology views biological function and macroscopic behavior as an emergent or supervenient property – i.e., a property that a collection of components or complex system possesses but which the individual constituents do not have. The properties of individual elements, such as proteins, are investigated in the context of the whole, complex system of interactions. The different spatial and temporal scales involved in biological processes – ranging from the level of molecules through to organisms and, ultimately, to the level of entire populations or ecosystems – permit upward and downward causation in complex arrangements of feedback loops. Systems-level properties arise from interconnected processes on multiple

scales of temporal and spatial organization. Understanding such complexity is a major challenge to the unaided human brain. Thus, using mathematical and computational models, systems biologists integrate elementary processes of systems into a coherent description that allows them to predict and characterize the systems-level properties and behavior of complex biological phenomena.

As the field of systems biology matures, we are beginning to see practical answers to real biological problems. We believe it is now time to step back and review some of the approaches of systems biology to concrete problems. This volume introduces some of the main methods and techniques of systems biology and assesses their pros and cons based on concrete case studies. The investigated biological phenomena include tissue organization, hormonal control, bacterial stress response, tumor growth and cellular metabolism. Each chapter and the book as a whole is intended to simultaneously serve as *design blueprint*, *user guide*, *research agenda*, and *communication platform*.

As *design blueprint*, the book is intended for biologists, mathematicians and systems scientists, computer scientists and technology developers, managers, and other professionals who consider adopting a systems biology approach.

As *user guide*, this volume addresses the requirements of scientists and researchers to gain an overview and a basic understanding of key systems biology methodologies and tools. For these users, we seek to explain the key concepts and assumptions of the various techniques, their conceptual and computational merits and limitations, and, where possible, give guidelines for choosing the methods and tools most appropriate to the task at hand. Our emphasis is not on a complete and intricate formal and technical treatment of the presented methodologies. Instead, we aim at providing the users with a clear understanding and practical know-how of the relevant methods in the context of concrete life science problems.

As *research agenda*, the book is intended for computer and life science students, teachers, researchers, and managers who seek to understand the state of the art of the methodologies used in systems biology research and development. To achieve this, we have attempted to cover a representative range of life science areas and systems biology methodologies, and we have asked the authors to identify areas in which gaps in our knowledge demand further research and development.

The book is also intended as a *communication platform* to bridge the cultural, conceptual, and technological gap among the key systems biology disciplines of biology, mathematics, and information technology. To support this goal, we have asked the contributors to adopt an approach that appeals to audiences from different backgrounds.

Providing a representative overview of current research, this book aims to illustrate the insights gained by adopting a systems biology approach. While systems biologists typically apply mathematical, statistical, and computational methods, these insights are presented in the context of current life science research. As a result, this book is targeted at an interdisciplinary audience comprising life scientists, mathematicians, system and computer researchers, and developers. In pursuing these goals, the book seeks to bridge the cultural, conceptual, and technological gap among the key disciplines that contribute to systems biology.

**Table 1** Classification of modeling formalisms: examples

|            | Deterministic                       | Stochastic             |
|------------|-------------------------------------|------------------------|
| Continuous | ODE, PDE                            | SDE                    |
| Discrete   | Boolean network, cellular automaton | Agent-based simulation |

In recent years, the increased interest of computer scientists in systems biology has led to an explosion of novel systems methodologies for modeling, analysis, and validation, but also for model representation and exchange. In this book, we do not intend to cover a wide variety of these methods, but we aim to present illustrative applications of systems biological methods in a representative overview.

In any modeling discipline, modeling formalisms may be classified according to the type of representation chosen to model *time*, *space*, and *entities* (such as the cell, proteins, or genes) of the system. These entities or dimensions can be modeled as continuous variables, so that the model can cope with any value within a meaningful range. Table 1 illustrates this. *Continuous* means that the model may output a simulation result at any given time point,  $t$  (continuous time) and location,  $x$  (continuous space), and that the output of the model may assume any value within a predefined range. In contrast, *discrete* refers to a modeling strategy that uses distinct values from a predefined set to represent time, space, and the entities of the modeled system. The output of a time-discrete model is limited to certain time intervals; a space-discrete model can explore only certain points in a given space; and discrete variables express levels or predefined states (*on/off*, *low/high*, cell cycle phase) of the modeled entities. Clearly, any of the combination of discrete and continuous methods is possible. An agent-based simulation can be backed by a time-continuous, space-discrete model with agents that are represented using both continuous and discrete variables. Discrete methodologies sometimes deviate from the classification shown in Table 1. The most common cases are shown in the table.

Systems biology modeling methodologies may also be divided into deterministic and stochastic formalisms. Consider a set of interacting cells which behave according to certain rules. In reality, the observation of randomly picked single cells may lead to grossly varying observations; although when looking at a large number of cells, they all share the same characteristic behavior. Deterministic simulations deal with this problem by modeling only those characteristics; the stochastic approach, in contrast, considers a large number of individual simulations and uses statistical analysis to draw conclusions.

Below we provide a brief overview of the contributed chapters in terms of the modeling methodology used and the biological problems addressed.

The modeling framework that was probably the first to be adapted for systems biological modeling – before the term systems biology was even coined – is the mathematical framework with the longest tradition: differential equations modeling, or more concrete, *ordinary differential equations (ODEs)*. The ODE methodology offers a variety of basic, mathematical, and computational tools for modeling, simulation, and qualitative and quantitative analysis.

Chapter 1 presents two elementary case studies that illustrate ODE-based model definition as well as timescale analysis and sensitivity analysis. These analysis methods can be used to extract biologically meaningful information from the model. In the study, the authors measure the efficiency of the simulated cell's protein-folding machinery under various conditions using timescale analysis.

While ODEs offer a general and flexible approach to modeling, this methodology relies on a qualitatively and quantitatively exact definition of the molecular network or system to be represented. Chapter 2 illustrates some of the most common mathematical tools in an ODE-based case study relating to folate metabolism.

Chapter 3 presents a delay differential equations (DDE) model of hormonal control of the menstrual cycle. This study demonstrates that it is sometimes more interesting to characterize the behavior of a system in relation to its inputs and parameters, than to just reproduce its outputs using concrete parameter values.

Pharmacokinetic models, most of which are ODE-based, have become an established tool in pharmacology. Such models have become an important tool in drug development to predict the fate of drugs or toxins taken in by the human body. Chapter 4 introduces this field and highlights the problem of investigating active transport phenomena.

The studies presented in Chaps. 3 and 4 rely on a reasonably well-established body of quantitative data. However, in the majority of cases, sufficient amounts of data are currently not available to systems biologists. The need to abstract from concrete sets of parameters has therefore led to the development of different modeling methods. *Piece-wise linear (PL) equations*, introduced in Chap. 6, are one example. Based on ODEs, they divide the entire parameter space into parts that share the same qualitative behavior. This behavior is approximated using only simple, linear equations, as opposed to the nonlinear equations that typically arise in complex ODE systems. This property makes PL models mathematically more tractable.

*Flux balance analysis (FBA)* is another useful tool in pharmacological applications of systems biology. An FBA model can predict metabolic activities (fluxes) under homeostatic conditions. Knowing the relevant metabolites and the stoichiometry of all reactions in the system is sufficient for performing such an analysis. FBA permits comprehensive studies of qualitative structural changes in the network, such as deletion of arbitrary genes throughout the genome. Chapter 5 presents an FBA case study concerned with the metabolism and pathogenicity of *Mycobacterium tuberculosis*. The overall goal of the effort is to systematically and efficiently design anti-tuberculosis drugs. Toward this goal, this chapter also illustrates how other techniques, besides FBA, can be used. The use of graph-theoretical techniques are illustrated for analyzing the protein–protein interaction networks, to gain insights about strategic hub proteins and possible of routes of information flow in triggering drug resistance. Boolean network modeling, another technique gaining popularity for studying biological systems, has been used for studying host–pathogen interactions, in this case leading to qualitative understanding of the complex interplay of the bacterial components with the human immune system.

Another modeling technique which is growing in popularity is the *agent-based model* (or individual-based model). Chapter 7 illustrates this methodology with an

application to the problem of bacterial antibiotic resistance. In this model, each cell is represented as an agent, which moves and interacts with other agents according to a defined set of rules. The agent paradigm is well suited to investigating the mechanisms of emergent spatial patterns. This is also discussed in Chap. 8, where an agent-based model is used to mimic the assembly of microtubules into the mitotic spindle at cell division.

Since different modeling methodologies are typically suited for different scales of time and space, it is an appealing proposition to build multi-scale models, where multiple modeling techniques applied to different aspects of the same biological problem integrate into a single, integrated model. The agent-based modeling approach permits the use of arbitrary modeling methods for defining the rule sets by which the agents are governed. This is illustrated in Chap. 9, where agents are used to model the behavior of epithelial tissue.

Finally, Chap. 10 uses an entirely different approach to investigate a problem in synthetic biology. In this discipline, biological molecules are used to engineer functional entities such as logic circuits. In this study, a domain-specific programming language helps to model and define the behavior of this engineered component.

Coleraine  
August, 2010

Werner Dubitzky  
Jenny Southgate  
Hendrik Fuß



# Contents

|          |                                                                                                             |            |
|----------|-------------------------------------------------------------------------------------------------------------|------------|
| <b>1</b> | <b>Effects of Protein Quality Control Machinery on Protein Homeostasis .....</b>                            | <b>1</b>   |
|          | Conner I. Sandefur and Santiago Schnell                                                                     |            |
| <b>2</b> | <b>Metabolic Network Dynamics: Properties and Principles .....</b>                                          | <b>19</b>  |
|          | Neema Jamshidi and Bernhard Ø. Palsson                                                                      |            |
| <b>3</b> | <b>A Deterministic, Mathematical Model for Hormonal Control of the Menstrual Cycle .....</b>                | <b>39</b>  |
|          | R. Drew Pasteur and James F. Selgrade                                                                       |            |
| <b>4</b> | <b>Modeling Transport Processes and Their Implications for Chemical Disposition and Action .....</b>        | <b>59</b>  |
|          | Nick Plant                                                                                                  |            |
| <b>5</b> | <b>Systems Biology of Tuberculosis: Insights for Drug Discovery .....</b>                                   | <b>83</b>  |
|          | Karthik Raman and Nagasuma Chandra                                                                          |            |
| <b>6</b> | <b>Qualitative Analysis of Genetic Regulatory Networks in Bacteria .....</b>                                | <b>111</b> |
|          | Valentina Baldazzi, Pedro T. Monteiro, Michel Page, Delphine Ropers, Johannes Geiselmann, and Hidde de Jong |            |
| <b>7</b> | <b>Modeling Antibiotic Resistance in Bacterial Colonies Using Agent-Based Approach .....</b>                | <b>131</b> |
|          | James T. Murphy and Ray Walshe                                                                              |            |
| <b>8</b> | <b>Modeling the Spatial Pattern Forming Modules in Mitotic Spindle Assembly .....</b>                       | <b>155</b> |
|          | Chaitanya A. Athale                                                                                         |            |

|           |                                                                                        |     |
|-----------|----------------------------------------------------------------------------------------|-----|
| <b>9</b>  | <b>Cell-Centred Modeling of Tissue Behaviour</b> .....                                 | 175 |
|           | Rod Smallwood                                                                          |     |
| <b>10</b> | <b>Interaction-Based Simulations for Integrative Spatial<br/>Systems Biology</b> ..... | 195 |
|           | Antoine Spicher, Olivier Michel, and Jean-Louis Giavitto                               |     |
|           | <b>Glossary</b> .....                                                                  | 233 |
|           | <b>Index</b> .....                                                                     | 237 |

# Contributors

**Chaitanya A. Athale** EMBL, Meyerhofstrasse 1, Heidelberg 69117, Germany  
and  
IISER Pune, IISER, Central Tower, Sai Trinity Building, Sutarwadi Road, Pashan,  
Pune 411021, India, [athale@embl.de](mailto:athale@embl.de); [cathale@iiserpune.ac.in](mailto:cathale@iiserpune.ac.in)

**Valentina Baldazzi** INRIA Grenoble – Rhône-Alpes, France,  
[Valentina.Baldazzi@inria.fr](mailto:Valentina.Baldazzi@inria.fr)

**Nagasuma Chandra** Bioinformatics Centre, Indian Institute of Science,  
Bangalore 560 012, India, [nchandra@serc.iisc.ernet.in](mailto:nchandra@serc.iisc.ernet.in)

**Hidde de Jong** INRIA Grenoble – Rhône-Alpes, France,  
[Hidde.de-Jong@inria.fr](mailto:Hidde.de-Jong@inria.fr)

**Johannes Geiselmann** Université Joseph Fourier, Grenoble, France,  
[hans.geiselmann@ujf-grenoble.fr](mailto:hans.geiselmann@ujf-grenoble.fr)

**Jean-Louis Giavitto** IBISC Lab, FRE 3190 CNRS, Université d'Évry  
& Genopole, 523 place des terrasses de l'agora, 91000 Évry, France,  
[giavitto@ibisc.univ-evry.fr](mailto:giavitto@ibisc.univ-evry.fr)

**Neema Jamshidi** Department of Bioengineering, University of California,  
San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0142, USA, [neema@ucsd.edu](mailto:neema@ucsd.edu)

**Olivier Michel** LACL – EA 4219 – Université de Paris 12, Paris EST 61 avenue  
du Général de Gaulle, 94010 Créteil Cedex, France,  
[olivier.michel@univ-paris12.fr](mailto:olivier.michel@univ-paris12.fr)

**Pedro T. Monteiro** INRIA Grenoble – Rhône-Alpes, France  
and  
IST/INESC-ID, 9 Rua Alves Redol, 1000-029 Lisbon, Portugal,  
[Pedro.Monteiro@inria.fr](mailto:Pedro.Monteiro@inria.fr)

**James T. Murphy** Centre for Scientific Computing and Complex Systems  
Modeling, School of Computing, Dublin City University, Dublin 9, Ireland,  
[jamurphy@computing.dcu.ie](mailto:jamurphy@computing.dcu.ie)

**Michel Page** INRIA Grenoble – Rhône-Alpes, France, [Michel.Page@inria.fr](mailto:Michel.Page@inria.fr)

**Bernhard Ø. Palsson** Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0142, USA, [palsson@ucsd.edu](mailto:palsson@ucsd.edu)

**R. Drew Pasteur** Department of Mathematics and Computer Science, The College of Wooster, 1189 Beall Ave., Wooster, OH 44691, USA, [rpasteur@wooster.edu](mailto:rpasteur@wooster.edu)

**Nick Plant** Centre for Toxicology, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK, [n.plant@surrey.ac.uk](mailto:n.plant@surrey.ac.uk)

**Karthik Raman** Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India, [k.raman@bioc.unizh.ch](mailto:k.raman@bioc.unizh.ch)

**Delphine Ropers** INRIA Grenoble – Rhône-Alpes, France, [Delphine.Ropers@inria.fr](mailto:Delphine.Ropers@inria.fr)

**Conner I. Sandefur** Center for Computational Medicine and Bioinformatics, University of Michigan, 2017 Palmer Commons, 100 Washtenaw Ave, Ann Arbor, MI 48105, USA, [sandefur@umich.edu](mailto:sandefur@umich.edu)

**Santiago Schnell** Department of Molecular and Integrative Physiology, Center for Computational Medicine and Bioinformatics, Brehm Center for Type 1 Diabetes Research and Analysis, University of Michigan, 2017 Palmer Commons, 100 Washtenaw Ave, Ann Arbor, MI 48105, USA, [schnells@umich.edu](mailto:schnells@umich.edu)

**James F. Selgrade** Department of Mathematics and Biomathematics Program, North Carolina State University, Raleigh, NC 27695-8205, USA, [selgrade@math.ncsu.edu](mailto:selgrade@math.ncsu.edu)

**Rod Smallwood** Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, UK, [r.smallwood@shef.ac.uk](mailto:r.smallwood@shef.ac.uk)

**Antoine Spicher** LACL – EA 4219 – Université de Paris 12, Paris EST – 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France, [antoine.spicher@univ-paris12.fr](mailto:antoine.spicher@univ-paris12.fr)

**Ray Walshe** Centre for Scientific Computing and Complex Systems Modeling, School of Computing, Dublin City University, Dublin 9, Ireland, [ray@computing.dcu.ie](mailto:ray@computing.dcu.ie)

# Chapter 1

## Effects of Protein Quality Control Machinery on Protein Homeostasis

Conner I. Sandefur and Santiago Schnell

### 1.1 Protein Folding is Catalyzed by a Complex Network of Reactions

A driving force of systems biology is the desire to understand the many interactions that compose the pathways within a cell. Systems biology is interested in the interactions and emergent properties that result from communication between different system components. Reducing a system (e.g., a cell) to its parts (e.g., individual genes and proteins) neglects component interaction and emergent properties. Building and investigating a complete interaction map provides insight into normal and diseased individuals that might not be found by traditional methods.

Much of traditional biology has the central dogma of molecular biology at its basis. This dogma states that DNA is transcribed into RNA which is translated into protein (Crick 1970), and has guided the study of individual genes and the proteins they encode. The protein folding network provides an example of how the central dogma of molecular biology does not explain many of the interactions within cells. DNA transcription is initiated by proteins and is the first step in protein production. For a number of eukaryotic proteins, the process continues with co-translation through ribosomes into the endoplasmic reticulum (ER). Molecular chaperones and folding machinery aid in folding protein into its native structure. This native state is not a random one but is instead the result of both the amino acid sequence and the complex folding network. These properly folded proteins are transported out of the ER for further processing.

The path from gene to protein is composed of many different and unknown interactions between DNA, RNA, proteins, and small molecules. Protein folding is one network, or subsystem, within the larger system of protein production. A systems biology approach offers us an opportunity to understand the complicated network of protein folding and the emergent properties that arise from interacting

---

C.I. Sandefur (✉)

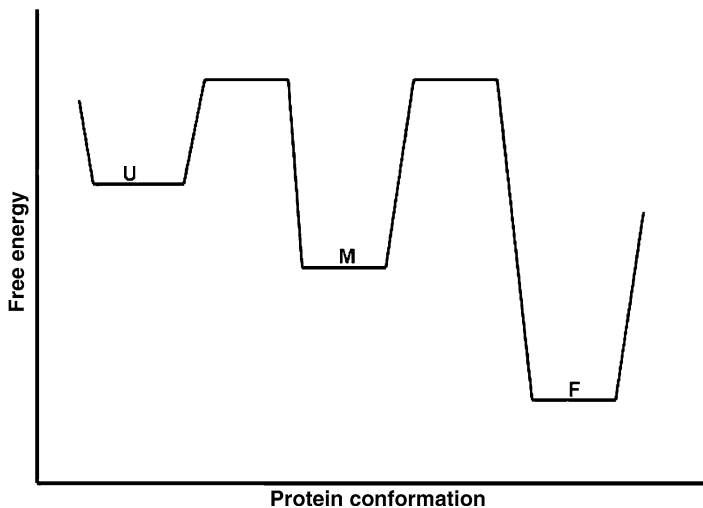
Center for Computational Medicine and Bioinformatics, University of Michigan,  
2017 Palmer Commons, 100 Washtenaw Ave, Ann Arbor, MI 48105, USA  
e-mail: [sandefur@umich.edu](mailto:sandefur@umich.edu)

network components. In this chapter, we explore two models of protein folding and misfolding to investigate how the protein folding network affects protein homeostasis. Using these models, we can identify the protein quality control pathways regulating folding and offer potential therapeutic targets for protein folding diseases.

### ***1.1.1 Disruptions to the Protein Folding Network are Associated with Disease***

Protein folding is often described by way of a folding energy landscape (Fig. 1.1) (Chiti and Dobson 2006). The landscape is composed of different conformations of a given protein each corresponding to a different energy level. The minimum energy, three-dimensional folded protein structure is termed the “native state” and for most proteins, is essential for proper function (Alberts et al. 2008). Failure to fold properly results in misfolded protein conformations. These protein conformations correspond to energy minima pockets within the folding energy landscape.

Proteins may fail to properly fold through mutations, cellular stress, or stochastic events (Nakatsukasa and Brodsky 2008). A breakdown in the quality of protein production can lead to the accumulation of toxic levels of misfolded and unfolded proteins. Improperly folded proteins can form aggregates (Morimoto 2008). When the level of aggregates reaches a certain concentration threshold, these protein complexes may lead to proteotoxicity.



**Fig. 1.1** *Different protein conformations have different energies.* While the goal is to reach the lowest energy as a properly folded protein (F), some misfolded proteins (M) are located in energy minima. Unfolded protein is denoted by U

A variety of diseases are linked to protein misfolding. For example, disruption of proinsulin folding in  $\beta$ -cells is sufficient to induce diabetes in both humans and mice (Scheuner and Kaufman 2008). Aggregation due to increased protein misfolding is implicated in the neurological diseases Alzheimer's, Parkinson's, and Huntington's (Soto 2003). The mechanisms behind aggregation of misfolded proteins and how the cell copes with misfolded protein accumulation are unknown.

### ***1.1.2 The ER Functions as a Protein Folding Factory***

Despite many technological advances, a complete understanding of the process of protein folding remains elusive. Proteins fold by transitioning through intermediates that comprise the folding landscape. However, detecting intermediate structures is difficult. This is because fast folding intermediates are not easily measured using current technology (Dobson 2004).

The ER is responsible for the synthesis, folding, assembly, and modification of one third of the eukaryotic proteome (Kaufman 2004). Most proteins cannot refold into their native states in the absence of cellular machinery. Protein folding in the ER is analogous to a factory assembly line with machinery processing proteins into a final, unique, native conformation. Enzymes and molecular chaperones are a part of this machinery working along the protein assembly line. Once a protein is properly folded, it is exported from the ER. If unfolded or misfolded proteins accumulate in the ER factory above a certain threshold, protein homeostasis is disrupted which can result in proteotoxicity (Ron and Walter 2007).

Cells have evolved a set of quality control processes that restore protein homeostasis. The processes are collectively termed the unfolded protein response (UPR). The UPR aids in quality control of protein production through three general processes. One process of the UPR prevents the influx of new peptides into the ER (Harding et al. 1999). Halting incoming materials into the factory reduces the burden on the cellular machinery.

The second process of the UPR increases the capacity of the ER-assisted-folding (ERAF) pathway through upregulation of chaperones and folding catalysts. This additional machinery aids in efficient processing of proteins within the burdened factory. Along with assisting in protein folding, chaperones and enzymes also sequester polypeptides within the ER. This is done to ensure that the mature folded proteins meet the factory quality control standards before export (Brodsky 2007).

Third, the UPR invokes the ER-assisted-degradation (ERAD) pathway. Due to the strict quality control measures of the protein factory, most proteins are near degradation as they move along the assembly line (Liberek et al. 2008). Chaperones escort proteins targeted for degradation. The chaperones prevent aggregation by allowing proteins to remain soluble and accessible to retrotranslocation machinery (Nakatsukasa and Brodsky 2008). After a protein is retrotranslocated to the cytosol by a retrotranslocon channel, it is degraded by the ubiquitin/proteasome pathway (Meusser et al. 2005). Enhancement of degradation reduces the assembly line load.

### ***1.1.3 Mathematical Models of Protein Quality Control Provide Novel Insights into the Regulation of Protein Assembly***

Although great strides have been made in understanding the network of protein folding, we lack a complete picture of the processes necessary for proteins to properly fold. We can apply modeling to investigate the mechanisms of protein quality control and make new experimental predictions. In biochemical processes, mathematical models are generally systems of ordinary equations. Using these models, we can investigate how varying reaction rates impact relative levels of system components through time. Also, we can obtain a dynamical view of the impact of protein quality control on the synthesis of native protein.

We know that protein folding in the ER involves a quality control mechanism, but how does this impact the dynamics of the native protein concentration? Experimental observations of protein quality control show it to be dependent on the amount of protein within the ER lumen (Ron and Walter 2007). We hypothesize that this dependence increases the timescale of protein accumulation and depletion under quality control. We test this hypothesis by comparing two models of protein folding, one without quality control and the other with.

## **1.2 Case Studies**

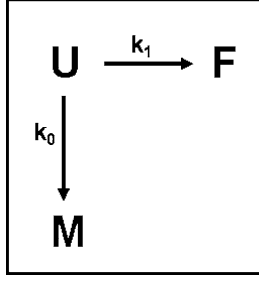
In the following case studies, we analyze two models of protein folding. The first case study is an analysis of a simple model describing protein folding in absence of the UPR. We follow with a second model describing protein folding regulated by the UPR. A comparison of the two models serves to illustrate how mathematical models provide a greater understanding of the dynamics of protein quality control.

### ***1.2.1 Case Study I: Protein Folding Without Quality Control***

The experimental measurements obtained from protein folding in vitro led to the development of the two state model of protein folding. In this model, unfolded protein spontaneously folds into its native state without intermediates (Anfinsen et al. 1954). This model provides a simple description of protein folding in absence of quality control machinery.

#### **1.2.1.1 Assumptions**

This first model contains three protein conformations: unfolded protein (U), folded protein (F), and misfolded protein (M) (Fig. 1.2). We are not considering influx or



**Fig. 1.2** *Schematic of protein folding without quality control.* The three protein conformations are represented as follows: unfolded (U), folded (F), and misfolded (M). Folding and misfolding reaction velocities are first-order with rate constants  $k_1$  and  $k_0$ , respectively. We assume that folding and misfolding are irreversible reactions. There is no influx or outflux of protein so the total protein concentration is conserved

outflux of protein; the system is closed and the total protein concentration is constant ( $u + m + f = \text{constant}$ ). Note that we denote protein concentrations using lower case variables.

We model spontaneous folding of unfolded protein at a rate of  $k_1$  and misfolding at a rate of  $k_0$  (Anfinsen et al. 1954). In general, chaperones are required for unfolding from a misfolded or folded state (Martin and Hartl 1997). Here, we assume that both folding and misfolding reactions are irreversible.

Equations (1.1)–(1.3) describe protein folding and misfolding in the absence of quality control by a linear system of ordinary differential equations.

$$\frac{du}{dt} = -(k_1 + k_0) u \quad (1.1)$$

$$\frac{dm}{dt} = k_0 u \quad (1.2)$$

$$\frac{df}{dt} = k_1 u. \quad (1.3)$$

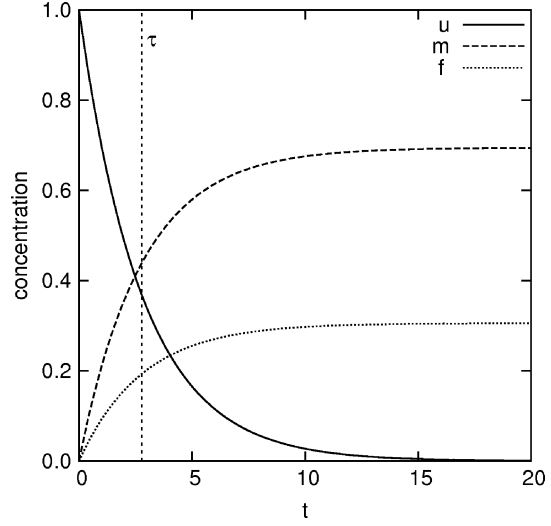
Note that the rate equations describing folded and misfolded protein are both dependent on unfolded protein.

### 1.2.1.2 Analytical Solution

We can solve this linear model analytically. Setting  $u_0$  as the total basal protein concentration ( $u(0) = u_0$ ,  $m(0) = 0$ , and  $f(0) = 0$ ), we find the analytical solution of our system to be:

$$u(t) = u_0 e^{-(k_0 + k_1)t} \quad (1.4)$$

**Fig. 1.3** Time course of the three protein conformation concentrations in absence of quality control. We begin with a basal unfolded protein ( $u$ ) concentration,  $u_0$ , of 1  $\mu$ M. Misfolded protein concentration,  $m$ , reaches a maximum of  $\frac{k_0}{k_0+k_1}u_0$  and folded protein concentration,  $f$ , reaches a maximum of  $\frac{k_1}{k_0+k_1}u_0$ . In this figure,  $k_0 = 0.25 \text{ s}^{-1}$  and  $k_1 = 1 \text{ s}^{-1}$ . The timescale for the system is denoted by  $\tau$



$$m(t) = u_0 \frac{k_0}{k_0 + k_1} \left( 1 - e^{-(k_0+k_1)t} \right) \quad (1.5)$$

$$f(t) = u_0 \frac{k_1}{k_0 + k_1} \left( 1 - e^{-(k_0+k_1)t} \right). \quad (1.6)$$

We can plot the concentrations of the different protein conformations as functions of time (Fig. 1.3). We begin with some basal unfolded protein concentration ( $u_0$ ) which decreases monotonically to zero. Misfolded protein levels increase towards a maximum misfolded concentration,  $m_{\max}$ , while folded protein levels increase towards a maximum folded protein concentration,  $f_{\max}$ , where,

$$m_{\max} = \frac{k_0}{k_0 + k_1} u_0 \quad \text{and} \quad (1.7)$$

$$f_{\max} = \frac{k_1}{k_0 + k_1} u_0. \quad (1.8)$$

### 1.2.1.3 Timescale Analysis

The timescale is the amount of time required for a significant change in the level of a protein conformation to occur and can be defined as (Segel 1984):

$$\text{timescale of } x(t) \approx \frac{x_{\max} - x_{\min}}{\left| \frac{dx}{dt} \right|_{\max}}. \quad (1.9)$$

Since the rates of formation of folded and misfolded protein depend on unfolded protein, the two terminal protein conformations are formed under the same timescale as unfolded protein depletion. The timescale of unfolded protein depletion and misfolded and folded protein accumulation is

$$\tau = \frac{1}{k_0 + k_1}. \quad (1.10)$$

In the initial transient of the folding process, the levels of misfolded and folded protein increase, as the misfolding and folding reactions compete for the unfolded protein (Fig. 1.3). Eventually, all of the unfolded protein in the system is either converted to folded or misfolded protein at rates  $k_1$  or  $k_0$ , respectively. If either rate is increased, unfolded protein is depleted from the system more quickly. If we increase the rate of folding,  $k_1$ , the maximum concentration of folded protein in the system increases. This also results in a decrease in the timescale of folded protein accumulation. We observe similar behavior in the misfolded protein levels when the rate of misfolding is increased.

#### 1.2.1.4 Conclusions for Case Study I

We introduced a simple model of protein folding and misfolding in absence of quality control. There is one timescale in the system that is dependent on the rates of folding and misfolding alone. Unfolded protein is depleted from the system on the same timescale as misfolded and folded protein form. In this linear system, the exact amounts of folded and misfolded protein can be determined at any time point by knowing the rates of misfolding and folding and the basal unfolded protein concentration. This model is a simplification and does not capture the interactions between the components of the cellular folding network in the ER. These interactions impact the overall behavior of the system as we will show in the next subsection.

### 1.2.2 Case Study II: Protein Folding with Quality Control

In Case Study I, we analyzed a model describing protein folding in the absence of the UPR. In reality, protein homeostasis within the folding factory of the ER is much more complicated. In this second case study, we analyze a model of protein folding regulated by the UPR. We compare the two models to investigate the impact of protein quality control machinery on protein homeostasis. We also perform a sensitivity analysis to identify parameters driving folding and misfolded protein accumulation. We discuss potential therapies for recovering folded protein levels under conditions promoting the accumulation of misfolded protein, such as those observed in protein misfolding diseases.

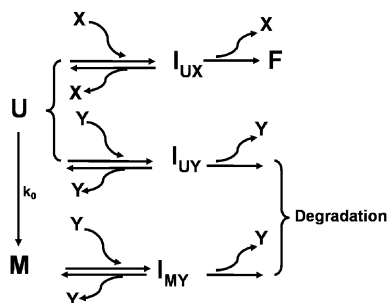
### 1.2.2.1 Assumptions

We analyzed a recently formulated model of the UPR in pancreatic  $\beta$ -cells (Fig. 1.4) (Schnell 2009). This model assesses factory function after activation of the three responses of the UPR.

As in Case Study I, we assume there is no input of unfolded protein into the system. Halted protein influx results in a reduction of protein entry into the ER lumen and is one of the three responses of the UPR (Harding et al. 1999). We begin with a basal unfolded protein concentration denoted  $u_0$ . We also assume that the rate of protein misfolding ( $k_0$ ) follows first-order kinetics and is proportional to the level of unfolded protein (Nolting 2006). Again, we model protein misfolding as irreversible (Martin and Hartl 1997).

It has been experimentally demonstrated that complex biochemical processes can be modeled as single enzyme reactions (Aldridge et al. 2006; Kholodenko 2006; Wiseman et al. 2007). Using this precedent of describing biochemical processes, two additional UPR processes were introduced into the quality control model of protein folding. As discussed above, ERAF and ERAD responses of the UPR are complex pathways comprised of many different components including chaperones and folding or degradation catalysts. Here, the ERAF response is modeled as a single enzyme with unfolded protein as a substrate (see Segel (1984) for details on modeling enzyme kinetics). The maximum velocity of folding is  $V_f$  with a Michelis–Menten (MM) constant of  $K_f$ . This MM constant is representative of the disassociation constant of folding machinery from unfolded protein.

Since a buildup of unfolded and misfolded protein in the ER lumen (which leads to an activation of the UPR) is assumed, the ERAD degradation machinery is modeled as responsible for removing both protein conformations (Nakatsukasa and Brodsky 2008). Therefore, a competition occurs between unfolded and misfolded



**Fig. 1.4** Schematic of protein folding with quality control. U is unfolded protein, F is folded protein, and M is misfolded protein. X is the enzyme representative of the folding machinery. Y is the enzyme representative of the degradation machinery. The enzyme–substrate complex intermediate for each pathway is represented by  $I_{UX}$ ,  $I_{UY}$ , and  $I_{MY}$ . Misfolding occurs through a first-order reaction with rate constant  $k_0$

protein as both are degraded with the same machinery. The maximum velocities of unfolded and misfolded protein degradation are denoted by  $V_u$  and  $V_m$ , respectively.  $K_u$  corresponds to the disassociation constant of degradation machinery from unfolded protein.  $K_m$  corresponds to the disassociation constant of degradation machinery from misfolded protein. Using the model schematic in Fig. 1.4 and the MM terms for the ERAF and ERAD process, we write the following system of differential equations describing protein folding under quality control:

$$\frac{du}{dt} = -k_0 u - \frac{V_f u}{K_f + u} - \frac{V_u u}{K_u \left(1 + \frac{m}{K_m}\right) + u} \quad (1.11)$$

$$\frac{dm}{dt} = k_0 u - \frac{V_m m}{K_m \left(1 + \frac{u}{K_u}\right) + m} \quad (1.12)$$

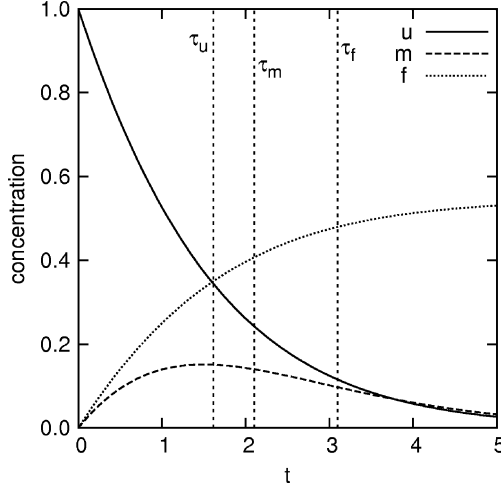
$$\frac{df}{dt} = \frac{V_f u}{K_f + u}. \quad (1.13)$$

### 1.2.2.2 Qualitative Dynamical Behavior and Equilibrium Points

Most nonlinear dynamical systems, such as the one described by (1.11)–(1.13), will not have an analytical solution. There are a variety of techniques useful for ascertaining the behavior of dynamical systems in this situation. In our analysis, we find the equilibrium points of the system (Sect. 1.2.2.2), we estimate the timescales (Sect. 1.2.2.3), and follow with a parametric sensitivity analysis to determine how the kinetic parameters impact the system (Sect. 1.2.2.4).

In order to find the equilibrium points of a system, we look for situations where all of the rate equations are equal to zero. In this system, the only equilibrium point is the trivial one:  $(u^*, m^*) = (0, 0)$ . Over time, all of the basal unfolded protein will either fold, misfold, or degrade (Fig. 1.5). Misfolded protein undergoes degradation as well, and therefore, both unfolded and misfolded protein concentrations are reduced to zero.

The minimum and maximum amounts of unfolded protein are the same across the two models. We expect the level of unfolded protein in both models to monotonically decrease from  $u_0$  to zero. The maximum misfolded protein concentration is different between the two models. In absence of quality control, the maximum amount of misfolded protein is only related to the rates of folding and misfolding. Under quality control, the misfolded protein reaches a maximum level due to misfolding but is also undergoing some level of degradation. The misfolded protein is eventually depleted to a zero concentration by degradation machinery.



**Fig. 1.5** The time course of the unfolded protein ( $U$ ), misfolded protein ( $M$ ) and folded protein ( $F$ ) concentrations under quality control. The timescales for unfolded protein depletion ( $\tau_u$ ), misfolded protein depletion ( $\tau_m$ ), and folded protein production ( $\tau_f$ ) are denoted by the vertical lines. The time course of degraded protein is not represented. The parameter values used were  $k_0 = 0.25 \text{ s}^{-1}$ ,  $V_f = V_m = 1.0 \text{ } \mu\text{M s}^{-1}$ ,  $V_u = 0.1 \text{ } \mu\text{M s}^{-1}$ ,  $K_f = 2.1 \text{ } \mu\text{M}$ , and  $K_u = K_m = 1.1 \text{ } \mu\text{M}$

### 1.2.2.3 Timescale Analysis

As in Case Study I, we determine the timescale for a process by estimating (1) the maximum and minimum concentrations of a given protein conformation and (2) the magnitude of the maximum reaction rate describing the evolution of the protein conformation over time. However, timescale determinations of non-linear systems is also a bit of an art. It requires making simplifying assumptions using our biological intuition about the system (Segel 1972; Segel and Slemrod 1989).

We begin by looking at the timescale for unfolded protein depletion. We know that the minimum amount of unfolded protein is 0 and the maximum is  $u_0$ . At the beginning of the reaction, the level of misfolded protein is small ( $m(t \approx 0) \approx 0$ ) while the level of unfolded protein is near the basal unfolded protein concentration ( $u(t \approx 0) \approx u_0$ ). We use this information to estimate  $\left| \frac{du}{dt} \right|_{\max}$  from (1.11) as:

$$\left| \frac{du}{dt} \right|_{\max} \approx u_0 \left( k_0 + \frac{V_f}{K_f + u_0} + \frac{V_u}{K_u + u_0} \right). \quad (1.14)$$

Applying (1.9) gives the timescale for the unfolded protein depletion:

$$\tau_u = \left( k_0 + \frac{V_f}{K_f + u_0} + \frac{V_u}{K_u + u_0} \right)^{-1}. \quad (1.15)$$

Misfolded protein initially accumulates, reaches a maximum level and then undergoes depletion. We can also determine the timescale for misfolded protein

depletion through some simplifications. We focus on the depletion phase as we are interested in understanding the behavior of misfolded protein under conditions of quality control. The depletion phase begins when the concentration of misfolded protein is at its maximum. We overestimate the maximum amount of misfolded protein as  $u_0$  for the depletion phase. The minimum value of misfolded protein is zero. During the depletion phase, we assume that the contribution of unfolded protein is negligible because it has been depleted due to folding, misfolding, or degradation, so we treat  $u(t) \approx 0$ . This allows us to approximate the maximum rate for misfolded protein from (1.12) as:

$$\left| \frac{dm}{dt} \right|_{\max} \approx \frac{V_m u_0}{K_m + u_0}. \quad (1.16)$$

We apply (1.9) to determine the timescale for the depletion phase of misfolded protein as shown:

$$\tau_m = \frac{K_m + u_0}{V_m}. \quad (1.17)$$

In order to restore homeostasis, the degradation machinery works to remove misfolded protein from the ER lumen. For example, we see from (1.17) that increasing the maximum velocity of misfolded degradation reduces the timescale and restores homeostasis more quickly.

We conclude with an estimation of the folded protein accumulation timescale. Here, we know that the minimum amount of folded protein is zero (the initial folded protein concentration). We overestimate the maximum folded protein concentration by allowing all of the basal unfolded protein to fold ( $f_{\max} = u_0$ ). By applying (1.9), we obtain the timescale for folded protein accumulation from (1.13) as:

$$\tau_f = \frac{K_f + u_0}{V_f}. \quad (1.18)$$

With these timescale estimations, we proceed with a comparison of two models to explore how protein quality control machinery impacts the dynamics of protein homeostasis. The timescale for unfolded protein depletion and misfolded protein accumulation in absence of quality control is dependent on the rate of misfolding. Further, the timescale in Case Study I is dependent on  $k_1$ . Adding protein quality control machinery essentially replaces  $k_1$  with  $\frac{V_f}{K_f + u_0} + \frac{V_u}{K_u + u_0}$ . We can set up the following inequality to explore the two timescales as shown:

$$k_1 \geq \frac{V_f}{K_f + u_0} + \frac{V_u}{K_u + u_0}. \quad (1.19)$$

When (1.19) holds, the timescale for unfolded protein depletion in Case Study I is shorter than the timescale of unfolded protein depletion under protein quality control. When (1.19) does not hold, the reverse is true.

We can reduce the timescale of unfolded or misfolded protein depletion under conditions of quality control by increasing the  $V_f$  or  $V_u$ , for example. By increasing

$V_f$  through increasing the amount of folding machinery, the timescale for unfolded protein depletion is reduced in this instance. Therefore, an example of a potential treatment for protein misfolding disease would entail targeting the maximum velocity of folding in order to restore protein homeostasis more quickly.

In absence of quality control, the folded protein accumulation timescale in (1.10) depends on the first-order misfolding and folding rates. Therefore, decreasing the misfolding rate, for example, would decrease the timescale for folded protein accumulation. It would also lead to an increase in the final, folded protein concentration in (1.8). In contrast, by our estimation in (1.19), the folded protein accumulation timescale under quality control does not depend on the misfolding rate. We cannot use the same argument of decreasing misfolding with the hopes of increasing folded protein levels when the system is under protein quality control.

In contrast to the timescale of Case Study I shown in (1.10), the timescales of the model with protein quality control are all dependent on the basal levels of unfolded protein in the system. Therefore, an increased amount of basal unfolded protein would lengthen the timescales of folded protein accumulation and unfolded and misfolded protein depletion under quality control. This is important since physiological changes can increase the demand for folded proteins leading to an influx of unfolded protein in the ER lumen. For example, increased blood glucose levels lead to an increased demand for insulin production in order to replace glucose-stimulated secreted insulin. The ER quality control must allow for proteins to fold while also managing the influx of new proteins. Cells could upregulate ERAD or ERAF machinery to manage this influx, but how would this impact folded protein levels? We use parametric sensitivity analysis to lend insight.

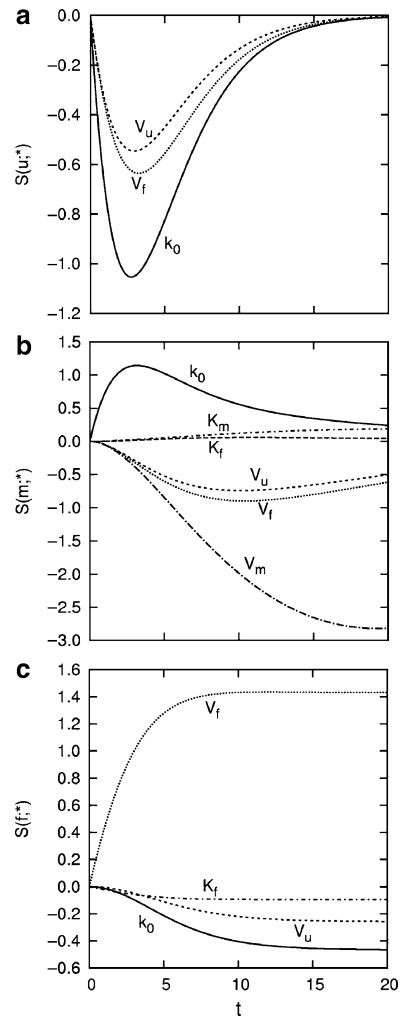
#### 1.2.2.4 Parametric Sensitivity Analysis

Parametric sensitivity analysis allows us to quickly ascertain how the parameters impact the system. We calculated the relative local sensitivities of each protein conformation to the system parameters (Varma et al. 1999). A sensitivity score is a way to quantify the relationship between system behavior and a parameter. When we calculate the parametric sensitivity, we are assessing how an input parameter impacts the evolution of the concentration of a protein conformation. The magnitude of the sensitivity score denotes the strength of the sensitivity. The larger the magnitude, the more sensitive the evolution of a protein conformation is to a parameter. If a parameter increases the accumulation of a protein conformation and diminishes depletion, the sign of the sensitivity score is positive. On the other hand, if a parameter increases the depletion of a protein conformation or diminishes accumulation, the sensitivity score is negative. If a protein conformation is not sensitive to a given parameter, the sensitivity score will be zero.

We begin by examining the sensitivity of the unfolded protein concentration for a given set of input parameters (Fig. 1.6a). All of the sensitivity scores are negative for unfolded protein. The unfolded protein concentration is most sensitive to  $k_0$ , followed by  $V_f$ , and third,  $V_u$ . All three rates reduce the amount of unfolded protein in the system. The  $k_0$  depletes unfolded protein through conversion to misfolded

**Fig. 1.6** Relative sensitivities of each protein conformation for the system parameters across time.

(a) unfolded protein concentration sensitivities,  $S(u;*)$ , (b) misfolded protein concentration sensitivities,  $S(m;*)$ , and (c) folded protein concentration sensitivities,  $S(f;*)$ . Parameters giving a relative sensitivity near to zero for a given protein conformation were not included in the graphs. Sensitivity scores are based on the following parameter values:  $k_0 = 0.25 \text{ s}^{-1}$ ,  $V_f = V_u = V_m = 0.1 \text{ } \mu\text{M s}^{-1}$ , and  $K_f = K_u = K_m = 1.1 \text{ } \mu\text{M}$



protein. The  $V_f$  depletes unfolded protein through conversion to folded protein by the folding machinery. The  $V_u$  depletes unfolded protein levels through degradation. The concentration of unfolded protein is increasingly sensitive to all three parameters in the initial transient of the reaction. The sensitivities decrease in magnitude as unfolded protein is converted to other protein conformations or is degraded.

Misfolded protein accumulates to a maximum concentration due to conversion of unfolded protein to misfolded protein by first-order misfolding (accumulation phase) and then decreases due to degradation through ERAD (depletion phase). The  $V_m$  is the parameter to which misfolded protein accumulation is most sensitive (Fig. 1.6b). This parameter drives misfolded protein degradation. The sensitivity score for  $V_m$  increases in magnitude as the misfolded protein concentration increases during the accumulation phase. The  $k_0$  increases misfolded accumulation in the first

several time units, but the sensitivity score reduces (as the amount of unfolded protein available to misfold is depleted). The maximum velocities of  $V_u$  and  $V_f$  both reduce misfolded protein accumulation as noted by the negative sensitivity value. Folding or degradation of unfolded protein both reduce the amount of unfolded protein available to misfold.

In both case studies, folded protein reaches a maximum concentration, but in the second model, folded protein accumulation is under quality control. The concentration of folded protein is most sensitive to  $V_f$ . Increasing  $V_f$  enhances folded protein accumulation (Fig. 1.6c). The  $k_0$  diminishes the accumulation of folded protein by reducing the amount of available unfolded protein through misfolding, as denoted by the negative sensitivity score.

The  $V_u$  reduces the amount of unfolded protein available to fold through degradation. Returning to the question of whether upregulating ERAD would impact folded protein accumulation, we find that it does. If cells increased ERAD machinery, thereby increasing degradation of both misfolded and unfolded protein, reduced folded protein accumulation would result. Therefore, upregulation of general ERAD machinery as therapy could ease a burdened ER lumen, but also have the unintentional impact of reducing folded protein assembled in the ER factory.

From the sensitivity analysis, we observe that with quality control, recovery of folded protein levels may occur in two different ways. One manner of recovery is through reducing the impact of parameters diminishing folded protein accumulation ( $k_0$  and  $V_u$ ). If we decrease  $k_0$  or  $V_u$ , we increase unfolded protein levels during the initial transient of the process. A lower  $k_0$  also reduces the accumulation of misfolded protein. We also find that the  $V_u$  reduces the accumulation of misfolded protein.

We can also increase accumulation of folded protein in the UPR model by targeting the  $V_f$ . Not only does this increase folded protein accumulation, it also diminishes misfolded protein accumulation. This is in agreement with conclusions made by Schnell (2009). Our parametric sensitivity analysis also agrees with the Schnell's assessments of investigating the ERAD pathway for therapeutic targets in misfolded diseases. Our analysis highlights the importance of investigating the ERAF pathway as well. Targeted therapy towards the ERAF pathway may allow for both increased folded and decreased misfolded protein levels in protein misfolding diseases.

### 1.2.2.5 Conclusions for Case Study II

The concentration of cellular protein can reach 350 mg/mL, and without quality control processes, toxic levels of aggregates may result (Dobson 2004). The UPR works to reduce the build up of unfolded and misfolded proteins within the ER through attenuating protein synthesis and enhancing protein folding (ERAF) and degradation (ERAD) of non-native proteins. We used a recently formulated model of the UPR to explore processes of quality control.

The timescales of the model with quality control were dependent on the concentration of basal unfolded protein in the system. We found that the timescale for

unfolded protein depletion was dependent on the parameters driving protein folding as well as unfolded protein degradation. The timescales for misfolded protein depletion and folded protein accumulation were approximated based on overestimations of protein concentrations.

We also discussed therapeutic strategies based on parametric sensitivity analysis. There are several ways to increase folded protein accumulation under protein quality control. We can increase folded protein accumulation through decreasing parameter values that diminish folded protein accumulation or through increasing parameter values that enhance folded protein accumulation. Each strategy impacts the unfolded and misfolded protein conformations differently.

This model is a simplification of the UPR in the ER lumen. An advantage of this simplification is that this model of protein quality control is relatively tractable. An important aspect of mathematical modeling is validating predicted system behavior. Predictions made through mathematical modeling only help our progress in science when they are validated by experimental results. For example, one prediction is that increasing  $V_u$  would result in reduced levels of misfolded protein as well as diminished folded protein accumulation. Recall that  $V_u$  represents the maximum velocity at which unfolded protein degradation may occur. One manner of increasing this velocity is to increase the amount of degradation machinery in the ER. Putative proteins involved in ERAD to target for overexpression are OS-9 and XTP3-B in mammals and YOS9 in yeast (Nakatsukasa and Brodsky 2008). The results of this overexpression experiment would then be used to further refine the UPR model. Collaboration with experimentalists is a vital piece in developing mathematical models that both describe realistic behavior and that can be utilized to make realistic predictions for potential therapies.

### 1.3 Lessons Learned

1. Mathematical models are analogous to experimental tools used to test hypotheses. Simple models of protein homeostasis can be represented with first-order reaction kinetics. The treatment of complex biochemical pathways as single enzyme reactions can be used to create a more complete picture of the dynamics of protein homeostasis in the ER. We can use these models to explore how protein quality control machinery regulates the dynamics of protein homeostasis.
2. In the absence of protein quality control machinery, protein folding and misfolding rates can be modeled as first-order rate constants. Under these circumstances, increasing the folding rate results in both increased folded protein levels and decreased misfolded protein levels. Decreasing the rate of misfolding in the first model produces similar results.
3. First-order models (Case Study I) can be solved analytically. In Case Study II, we estimated timescales for a non-linear system. Scaling of non-linear systems is a bit of an art. Analysis of these models requires simplifications made using the biological intuition of the modeler (Segel 1972; Segel and Slemrod 1989).

4. Incorporating quality control machinery into a model of protein homeostasis impacts the overall behavior of the system. In the absence of quality control, one timescale describes the model and is based on the rates of misfolding and folding alone. By contrast, the timescales of the quality control model differed across the protein conformations. Also, the timescales were dependent on basal unfolded protein concentration as well as kinetic parameters describing protein misfolding, folding, and degradation.
5. A parametric sensitivity analysis identified how each protein conformation was impacted by the parameters in the model with quality control. According to our analysis, two different pathways can be targeted for therapeutic purposes. One potential therapy to improve folded protein levels involves reducing ERAD machinery to allow more unfolded protein to fold. Applying this therapy, could lead to proteotoxicity by allowing an accumulation of unfolded and misfolded protein within the ER lumen. The other potential therapy would be to target the ERAF pathway using pharmacological chaperones. Using this latter method, we decrease the accumulation of misfolded protein as well.

**Acknowledgements** The authors would like to acknowledge the comments from Marnie Briceno (University of Washington), Hannah Briolat (University of Michigan), and Michelle Wynn (University of Michigan). This work is based upon research supported by the National Science Foundation under Grant No. IIS-0852734.

## Appendix: Symbols Used in this Chapter

|          |                                                                 |
|----------|-----------------------------------------------------------------|
| $u(t)$   | Unfolded protein concentration at time $t$                      |
| $m(t)$   | Misfolded protein concentration at time $t$                     |
| $f(t)$   | Folded protein concentration at time $t$                        |
| $X$      | Folding machinery                                               |
| $Y$      | Degradation machinery                                           |
| $I_{UX}$ | Unfolded protein – folding machinery complex                    |
| $I_{UY}$ | Unfolded protein – degradation machinery complex                |
| $I_{MY}$ | Misfolded protein – degradation machinery complex               |
| $k_0$    | First-order misfolding rate                                     |
| $k_1$    | First-order folding rate                                        |
| $V_f$    | Maximum velocity of folding                                     |
| $K_f$    | MM constant of folding machinery-unfolded protein reaction      |
| $V_u$    | Maximum velocity of unfolded protein degradation                |
| $K_u$    | MM constant of degradation machinery-unfolded protein reaction  |
| $V_m$    | Maximum velocity of misfolded protein degradation               |
| $K_m$    | MM constant of degradation machinery-misfolded protein reaction |
| MM       | Michaelis–Menten                                                |

## References

- B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the cell*, 5th edition. Garland Science, San Francisco, 2008
- B. B. Aldridge, J. M. Burke, D. A. Lauffenburger, and P. K. Sorger. Physicochemical modeling of cell signalling pathways. *Nature Cell Biology*, 8:1195–1203, 2006
- C. B. Anfinsen, R. R. Redfield, W. L. Choate, J. Page, and W. R. Carroll. Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *Journal of Biological Chemistry*, 207:201–210, 1954
- J. L. Brodsky. The protective and destructive roles played by molecular chaperones during ERAD (endoplasmic-reticulum-associated degradation). *Biochemical Journal*, 404:353–363, 2007
- F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75:333–366, 2006
- F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970
- C. M. Dobson. Experimental investigation of protein folding and misfolding. *Methods*, 34:4–14, 2004
- H. P. Harding, Y. Zhang, and D. Ron. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature*, 397:271–274, 1999
- R. J. Kaufman. Regulation of mRNA translation by protein folding in the endoplasmic reticulum. *Trends in Biochemical Sciences*, 29:152–158, 2004
- B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology*, 7:165–176, 2006
- K. Liberek, A. Lewandowska, and S. Zietkiewicz. Chaperones in control of protein disaggregation. *The EMBO Journal*, 27:328–335, 2008
- J. Martin and F. U. Hartl. Chaperone-assisted protein folding. *Current Opinion in Structural Biology*, 7:41–52, 1997
- B. Meusser, C. Hirsch, E. Jarosch, and T. Sommer. ERAD: the long road to destruction. *Nature Cell Biology*, 7:766–772, 2005
- R. I. Morimoto. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes and Development*, 22:1427–1438, 2008
- K. Nakatsukasa and J. L. Brodsky. The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. *Traffic*, 9:861–870, 2008
- B. Nolting. *Protein folding kinetics – Biophysical methods (2nd edition)*. Springer, Berlin, 2006
- D. Ron and P. Walter. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Reviews Molecular Cell Biology*, 8:519–529, 2007
- D. Scheuner and R. J. Kaufman. The unfolded protein response: a pathway that links insulin demand with  $\beta$ -cell failure and diabetes. *Endocrine Reviews*, 29:317–333, 2008
- S. Schnell. A model of the unfolded protein response: pancreatic  $\beta$ -cell as a case study. *Cellular Physiology and Biochemistry*, 23:233–244, 2009
- L. A. Segel. *Modeling dynamic phenomena in molecular and cellular biology*. Cambridge University Press, Cambridge, 1984
- L. A. Segel. Simplification and scaling. *SIAM Review*, 14:547–571, 1972
- L. A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, 31:446–477, 1989
- C. Soto. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience*, 4:49–60, 2003
- A. Varma, M. Morbidelli, and H. Wu. *Parametric sensitivity in chemical systems*. Cambridge University Press, Cambridge, 1999
- R. L. Wiseman, E. T. Powers, J. N. Buxbaum, J. W. Kelly, and W. E. Balch. An adaptable standard for protein export from the endoplasmic reticulum. *Cell*, 131:809–821, 2007

# Chapter 2

## Metabolic Network Dynamics: Properties and Principles

Neema Jamshidi and Bernhard Ø. Palsson

### 2.1 Introduction

Dynamic descriptions of biological processes, especially metabolism, have been of interest for many years (Segel 1975). The size and complexity of these models, however, have stagnated for the last 20 years or so, in spite of dramatic improvements in computational capabilities. The development of large-scale kinetic models (hundreds to thousands of dynamic variables) has been deemed infeasible now for a number of years. Traditional approaches for parameterization of kinetic models require time and labor intensive biochemical assays on individual enzymes. This presents a challenge and a practical limitation to the number of enzymes that can be described using kinetic rate expressions and hence limits the size of networks that can be described dynamically. Furthermore, the confines of a microtiter plate or test tube are often significantly different than those of the intracellular environment. Hence, even once these measurements are carried out, they may not be relevant since the conditions were so different than the *in vivo* environment. Thus the development of genome-scale kinetic models with this approach has been recognized as infeasible. However, biology is a technology-driven science and new technologies have driven the understanding of biology through the ability to make deeper and broader measurements (e.g., fluxomics and metabolomics). Thus, these new data should analogously motivate and drive the development of new computational approaches.

Future development of network dynamics in biology, particularly with metabolism, will involve two branches, the construction of dynamic networks and the subsequent analysis and simulations of the resulting networks. This chapter will focus on the aspects of the latter; however, the first half will concern basic properties and features of dynamic networks, which will be relevant for both construction and analysis of networks. There are a number of reasons for interest in kinetic models: (1) the ability to make predictions about fluxes as well as concentrations, (2)

---

N. Jamshidi (✉)

Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive,  
La Jolla, CA 92093-0142, USA

e-mail: [neema@ucsd.edu](mailto:neema@ucsd.edu)

a more direct tie-in with experimental measurements, and (3) the ability to make a more direct connection with environmental as well as genetic perturbations by modifying the initial conditions and catalytic and binding constants. The analysis of kinetic models is a necessarily mathematical and computational topic, and there is often a tendency to lose the forest for the trees. The first few sections of this chapter will aim to place the dynamics in a broader context, so that it can be seen how it relates to steady-state flux-based models. While the principles and equations described herein will be applicable to most biological networks, our focus will be metabolism. Furthermore, we will focus on the dynamic hierarchy of metabolic networks with the aim of try to dissect and understanding the interactions that occur between components on different time scales.

## 2.2 Dynamic Mass Balances and Fundamental Subspaces

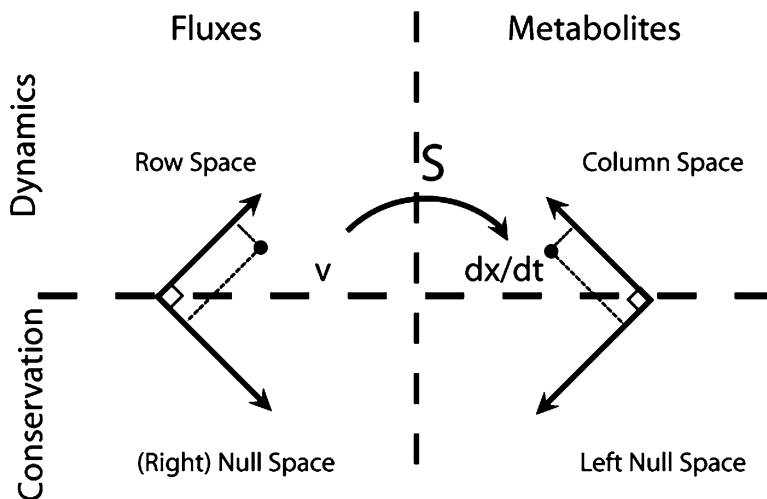
Representing biological interactions in terms of mathematical expressions enables one to be precise and unambiguous about what is being discussed. More importantly, however, this enables the ability to benefit from underlying mathematical properties reflected in the equations and the ability to apply physical constraints. For example, through mathematical representation of a metabolic network, one can enforce mass and energy conservation and then explore the implications of these constraints.

The dynamic mass balance equations that describe the dynamic states of biochemical reaction networks are (Heinrich et al. 1977; Reich and Selkov 1981):

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{S} \cdot \mathbf{v}(\mathbf{x}, \mathbf{k}) \quad (2.1)$$

in which  $\mathbf{x}$  is an  $m$ -dimensional vector of concentrations of the metabolites in the network in  $R^m$ ,  $\mathbf{v}$  is an  $n$ -dimensional vector of reaction fluxes in  $R^n$ ,  $\mathbf{k}$  represents a set of rate parameters, and  $\mathbf{S}$  is the  $m \times n$  stoichiometric matrix, containing the stoichiometric coefficients of reactants and products for each reaction in the network (Palsson 2006). The stoichiometric matrix,  $\mathbf{S}$ , is a mathematical representation of a metabolic pathway or network. Each column in the matrix corresponds to an enzymatic (or nonenzymatic) biochemical conversion which may be reversible or irreversible. The reaction flux vector contains rate expressions for all of the biochemical conversions described by  $\mathbf{S}$ .

If we consider a nonzero vector, there can be two types of results when the vector is multiplied by a matrix: it can result in a null vector (vector with 0 in all of the entries) or it can be nonzero. Vectors that yield null vectors when multiplied by a matrix lie in the null space. Pre-multiplication (multiplication from the left-hand side) of any column vector is actually a mapping from the column to the row space. Figure 2.1 pictorially illustrates what is described by (2.1). The four subspaces in Fig. 2.1 can be viewed as a  $2 \times 2$  table with each quadrant defined by fluxes or concentrations across the top and dynamics or conservation quantities on the side.



**Fig. 2.1** *The fundamental linear sub-spaces and their metabolic network interpretations.* The stoichiometric matrix,  $S$ , maps  $v$  which resides in the row and null spaces to  $dx/dt$ , which resides in the column and left null spaces. The null spaces describe conserved quantities (Palsson 2006); in the right null space this corresponds to conservation of flux and in the left null space this refers to conserved moieties within the network. The row and column spaces describe dynamic states. The row space/right null space and column space/left null space are orthogonal complement pairs, respectively (Strang 1988)

A complete study of the system properties of (2.1) would result in the characterization of all four subspaces of  $S$  (Strang 1988). The right null and left null spaces of  $S$  have been studied extensively over the past decades (Palsson 2006; Heinrich and Schuster 1996; Famili and Palsson 2003). The left lower box in Fig. 2.1, for example, is the set of flux balances reflecting mass conservation (total mass accumulated = total mass entering the system – total mass exiting the system). The bounds of the right null space confine the complete set of allowable steady-state flux distributions by enforcing the principle of mass conservation. This subspace has proven to be extremely insightful from a biological standpoint and has been studied extensively during the past decades (Palsson 2006; Heinrich and Schuster 1996). The left null space contains the time-invariant pools, which reflect conserved moieties or functional groups of metabolites in a particular system. Although there has been relatively less investigation into the properties of this subspace in metabolic networks, its significance and meaning has been well described (Palsson 2006; Famili and Palsson 2003).

The row and column spaces are the orthogonal complements to the null spaces, and while dynamic simulations of metabolism have been carried out since the very earliest days of the field of biochemistry (Segel 1975), dynamics are rarely discussed in terms of the subspaces and in relation to their orthogonal complements. Truly appreciating the general principles and underlying factors of dynamics requires recognition of their role and relationship to different subspaces. For example,

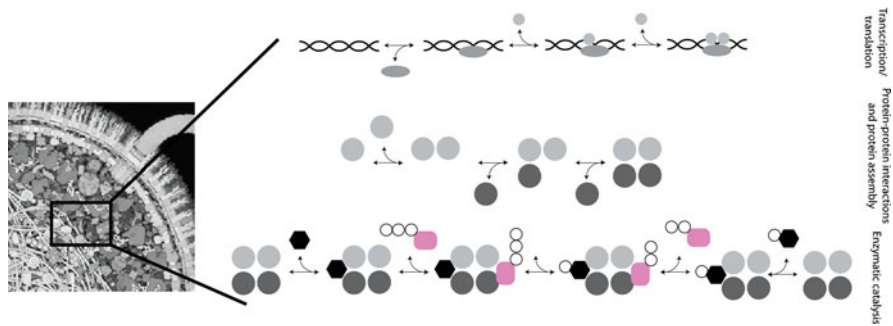
thinking about dynamics in terms of the row and column spaces (see Fig. 2.1) leads one to immediately recognize that the null spaces describe “conserved” quantities (i.e., mass) and the row and column spaces of  $\mathbf{S}$  describe the driving forces and the direction of motion for the network variables, respectively.

The topics and challenges of this chapter will focus on the top row of Fig. 2.1; however, occasional mention to the bottom row will be made, because these subspaces are not independent of one another. For example, if the total NAD moiety in a network is assumed to be constant (which would be identifiable through analysis of the left null space), there would be obvious implications in the analysis of the column space for NAD and NADH (knowing the dynamics of one would immediately inform the dynamics of the other).

### 2.2.1 Key Considerations in Networks

In the spirit of appreciating the components needed to construct dynamic networks, it is important to be cognizant of the nature of molecular interactions as well as some basic assumptions that are regularly made when building kinetic models.

The Michaelis–Menten rate equation is perhaps the most famous and commonly used rate expression for describing reaction kinetics. The rate expressions have proven to be extremely useful when the underlying assumptions have not been violated; however, in vivo conditions such as completely saturated enzymes are not always met. Hence, these rate expressions cannot be claimed to be valid in general. More to the point, the interactions that occur in biological networks, including macromolecular interactions and enzymatic catalysis, are all fundamentally bilinear interactions (Fig. 2.2). That is, most reactions involve two molecules combining to form a third. The general rate law for any of these steps is given by,  $v = kx_1x_2$ , in which  $k$  is a bilinear rate constant, and  $x_1$  and  $x_2$  are the concentrations of the



**Fig. 2.2** Molecular interactions are almost always combinations of bilinear association or dissociations

interacting components. Fortunately, due to increased processing speed and memory in computers, it is possible to begin describing large networks with complex regulatory schemes in terms of their bilinear interactions.

There is a long history of investigations into the dynamics and kinetics of metabolism. During this course, various mathematically driven operations and procedures have been developed. However, since biology is a technology-driven field, theories and formalisms are only as useful as their ability to integrate available data and to make testable predictions. The approach and formalism here, focusing on the **S** and **G** matrices, are predicated on capturing the key biological features of systems while also enabling the integration of available data types. Along these lines, the gradient matrix has not been defined in any way; it arises naturally from the linearization of the flux vector comprised of net elementary reaction rates. The gradient matrix describes the responses of the reactions to be changed in the concentrations. As will be described later, it is through **G** that the dual nature of the relationship between fluxes and concentrations can be developed.

The metabolic dynamics described by (2.1) also assumes that concentrations can be meaningfully defined (i.e., the number of compounds within the specified volume) and in the absence of any spatial gradients. There are clearly examples when these assumptions fail to be satisfied, for example, in transcription of genes when stochastic effects take place or with excitation–contraction coupling between muscle contraction and energy metabolism, when temporo-spatial gradients have significant effects. We will not address these issues here, but only caution that the modeler should take heed of the physicochemical environment of the phenomena that being modeled and to be cognizant of when particular assumptions may or may not be appropriate.

### 2.2.2 *Properties of Dynamic Systems*

Unfortunately, network dynamics are often discussed and viewed with a sense of “magic”, and an implication that somehow nonlinearity can make something appear out of nothing. However, if one understands the parts of a model and how they fit together, the results and predictions will be much more palatable and lead to an improved understanding of a network model and its behavior rather than increased confusion. We discuss some key properties of dynamic systems and how they contribute to the properties of networks. There are three matrices that will be of interest in this chapter: the stoichiometric, gradient, and Jacobian matrices. The stoichiometric matrix is a mathematical representation of the “links and nodes” of a network. The columns correspond to the links (or reactions) and the rows correspond to nodes (or chemical species/metabolites). The gradient matrix represents the dependence of the links on the nodes (in the linear regime). The Jacobian matrix is used to describe the overall dynamic relationships in the network; as will be seen however, this matrix can be composed from the stoichiometric and gradient matrices. The key point here is that the fundamental matrices of interest are the stoichiometric and gradient matrices, and these are in fact biological data matrices.

### 2.2.2.1 Underlying Structure of the Jacobian

Linearization of (2.1) as described in Sect. 2.6.1 results in the ability to define the Jacobian matrix as a product of the stoichiometric and gradient matrix,

$$\mathbf{J} = \mathbf{S} \cdot \mathbf{G} \quad (2.2)$$

The gradient matrix can then be factored, such that

$$\mathbf{J} = \mathbf{S} \cdot \mathbf{K} \cdot \mathbf{M} \quad (2.3)$$

in which  $\mathbf{K}$  is an  $n \times n$  diagonal matrix whose entries are the lengths of the rows of  $\mathbf{G}$  with units 1/time.

Hence, these entries are pseudo time constants or characteristic times corresponding to the reactions in the network. Consequently, the rows in  $\mathbf{G}$  indicate the direction that each reaction lies. So the  $\mathbf{K}$  and  $\mathbf{M}$  matrices describe kinetic and thermodynamic driving factors in the network.<sup>1</sup> This decomposition into the kinetic and thermodynamic influences was carried out without any involved mathematical procedures and has been determined by matrices with biologically meaningful interpretations.

### 2.2.2.2 Structural Similarity

Reaction rates are commonly expressed as the net sum of elementary reactions. When this is done, it follows that  $\mathbf{S}$  and  $\mathbf{G}^T$  are structurally similar (Jamshidi and Palsson 2008a) with the corresponding row and column entries have zero or nonzero values. This similarity underlies the stoichiometric influence in network dynamics. In spite of these similarities, there are also key differences between these matrices, which will be touched up on in Sect. 2.4.

### 2.2.2.3 Flux-Concentration Duality

The first property leads to the ability to define a pair of dual Jacobian matrices. One for the concentrations,

$$\mathbf{J}^x = \mathbf{S} \cdot \mathbf{G} \quad (2.4)$$

and one for the fluxes,

$$\mathbf{J}^v = \mathbf{G} \cdot \mathbf{S} \quad (2.5)$$

---

<sup>1</sup> Note that  $\mathbf{M}$  does *not* refer to the modal matrix in this chapter.

The systems described by each of these equations is the same; however, the independent variables is different, in one the variables are the concentrations and in the other the variables are the fluxes. Note that to convert from the concentration Jacobian to the flux Jacobian, not advanced mathematics or decompositions were used, simply reversing the order of multiplication of two matrices.

#### 2.2.2.4 Hierarchical Dynamics

A key feature of biological networks is the presence of many interactions that occur on a wide range of different time scales. Analysis of these properties has been active for many decades, and there is a rich history in time scale separation and modal decomposition of metabolic networks (Heinrich et al. 1977; Palsson and Lightfoot 1984; Okino and Mavrovouniotis 1998). This will not be detailed here, suffice it to say that one approach that has been successfully carried out for biochemical network analysis has been modal decomposition, which involves the diagonalization of the Jacobian matrix, and the redefinition of concentration variables into “modal” variables which move on dynamically independent time scales.

One overall goal of dynamic analyses of networks is to simplify network structure and to determine which interactions are relevant at particular time scales of interest. This enables one to filter out interactions that are either too fast or too slow to be of interest and to also characterize the progressive pooling of metabolites across slower and slower time scales.

### 2.3 Dual Jacobian Matrices

One consequence of recognizing the gradient matrix is that it leads to the definition of dual Jacobian matrices and highlights the nature of the relationship between concentration and flux dynamics. The duality between fluxes and concentrations results from the ability to define flux and concentration Jacobian matrices, as mentioned above. The nature of this relationship while mathematically interesting is also of importance from the biological perspective, and hence we will spend some additional time discussing the nature of the relationship. Measurements and perturbations are carried out in terms of concentration variables; however, analysis of the fluxes is what enables interrogation of the systemic properties of networks.

Each network only has a single stoichiometric matrix,  $\mathbf{S}$  and gradient matrix,  $\mathbf{G}$ . However, biological networks can be analyzed in terms of compound (node) variables or in terms of flux (link) variables. Thus, there are two Jacobian matrices describing the same network,  $\mathbf{J}^x = \mathbf{S} \cdot \mathbf{G}$  and  $\mathbf{J}^v = \mathbf{G} \cdot \mathbf{S}$ , depending on which variables, concentrations or fluxes, are used as state variables. The former gives a reaction-centric view of the dynamics, while the latter gives a compound-centric view. These are complementary views of the same system. The relevance of network topology in dynamic systems is highlighted by the fact that the Jacobian matrices

are weighted adjacency matrices containing weighted inner products of the reaction rows and columns ( $\mathbf{J}^v$ ) and compound rows and columns ( $\mathbf{J}^x$ ). Thus,  $\mathbf{J}^v$  and  $\mathbf{J}^x$  are structurally similar to the reaction adjacency matrix and the metabolite adjacency matrix, respectively (Palsson 2006).

Modal decomposition of the Jacobian has been previously applied for the analysis of biological networks. We note that the two Jacobian matrices share the same eigenvalues. The eigenvectors/rows  $\mathbf{J}^x$  relate to pool formation on various time scales (Heinrich et al. 1977; Palsson and Lightfoot 1984), while the eigenvectors/rows of  $\mathbf{J}^v$  relate the formation of groups of fluxes that move these pools (Jamshidi and Palsson 2008a). The key point to appreciate is that both views describe the same set of network interactions, but in terms of different dynamic variables; dynamic concentration variables in one case and dynamic flux variables in the other.

## 2.4 Stoichiometry Versus Gradients

Having stepped through the construction and deconstruction of biological networks, it is hopefully sufficiently impressed upon the reader that network dynamics can be comprehensively characterized through the definition of two matrices: the stoichiometric matrix,  $\mathbf{S}$ , and the gradient matrix,  $\mathbf{G}$ . This is a bold statement, and thus it will be followed by a bold caveat. The gradient matrix is rarely known in general for any condition; hence, experimental and measurement limitations require that it be characterized under a limited set of condition(s). Thus, one generally will only approximate the elements in the matrix and often be restricted to a linearized region close to a particular steady state.

It is important to recognize that  $\mathbf{S}$  and  $\mathbf{G}$  are data matrices, and they are not just of theoretical relevance but have very practical significance and import for the construction and subsequent analysis of kinetic networks (Jamshidi and Palsson 2008a). As mentioned above, when a network is described in terms of bilinear net elementary reactions, which are in general the most appropriate expressions,  $\mathbf{S}$  and  $\mathbf{G}^T$  have similar structures. In spite of these structural similarities, there are many important differences between these matrices, which we mention briefly here.

The stoichiometric matrix describes the chemical transformations and interconversions that occur among compounds in a network, and it is through the  $\mathbf{S}$  matrix that mass conservation can be enforced. The gradient matrix, on the other hand, accounts for the kinetic interactions that occur within a network and is constrained by thermodynamic bounds. With these differing physical constraints, there is subsequently different data types that are used to populate these matrices. Genomic and bibliomic data are needed to construct  $\mathbf{S}$  matrices. Alternatively, metabolomic, fluxomic, thermodynamic (e.g., equilibrium constants), and if possible kinetic data are needed to define  $\mathbf{G}$ .

The stoichiometric matrix contains integer entries; hence, it is a “knowable” matrix with the potential of no error associated with its elements. In contrast, the

elements of the gradient matrix are non-integer values and are subject to often significant experimental errors; hence, these entries may often only be known to an order of magnitude. The values within the gradient matrix may differ by more than 10 orders of magnitude; hence,  $\mathbf{G}$  is ill-conditioned and this underlies the stiffness of biological models, which may lead to difficulties when integrating the set of differential equations. However, it is also this wide range of values that leads to the characteristic time-scale separation in biological networks.

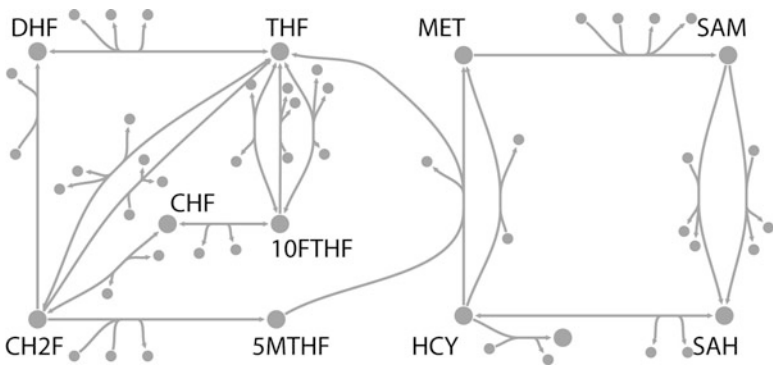
Thus, given these similarities and differences between  $\mathbf{S}$  and  $\mathbf{G}$ , the resulting biological interpretation of the matrices also differs. Each  $\mathbf{S}$  matrix is effectively a genomic representation of a species. Thus, different species will have different stoichiometric matrices, and changes in the  $\mathbf{S}$  results from “distal causation” (Mayr 1961). Conversely, the gradient matrix reflects the genetic features of individuals, and so although a species has a single  $\mathbf{S}$  matrix, a population will have a large set of differing  $\mathbf{G}$  matrices. Thus, the gradient matrix represents individual differences within populations and results from changes in “proximal causation” (Mayr 1961).

## 2.5 Example: Folate Metabolism

The remainder of this chapter will investigate some of the dynamic properties in a dynamic folate metabolic network. As mentioned above, there are many different avenues of analysis to focus on in dynamic networks, and the focus here will be on the metabolite pooling structure within the network on progressively slower time scales and the effects of environmental perturbations on this pooling. Reed et al. (2006) carried out multiple studies with folate one-carbon metabolism in humans with interesting observations with respect to nutrition and genetic variation. Since this is a validated yet small and relatively simple kinetic model, we focus on this network to discuss and highlight some of the discussion points earlier in this chapter.

### 2.5.1 *Constituent Matrices and Subspaces*

This network is described by 10 dynamic concentration variables and 20 reaction fluxes. As reflected by the network map (Fig. 2.3), there are often multiple interconversions between the same metabolites. The stoichiometric matrix and the gradient matrix appear in Tables 2.1 and 2.2. Note that all of the entries in  $\mathbf{S}$  are integers, whereas almost all of the entries in  $\mathbf{G}$  are real numbers. Since the model was not constructed strictly from mass action kinetics and various assumptions were made (e.g., Michaelis–Menten kinetics), the transpose of  $\mathbf{G}$  and  $\mathbf{S}$  is not similar in this case (however, if the relationships were explicitly described using mass action kinetics, the similarity between the two matrices would be preserved). Also note that the methionine input flux (metin) is zero for all of the entries in the gradient matrix.



**Fig. 2.3** *A map of the folate one-carbon metabolism network.* A map of the folate one-carbon metabolism network for the model described by [Reed et al. \(2006\)](#). Only the dynamic metabolite variables have been labeled. Abbreviations: *5MTHF* 5-methyltetrahydrofolate, *THF* tetrahydrofolate, *DHF* dihydrofolate, *CH2F* 5,10-methylenetetrahydrofolate, *CHF* 5,10-methenyltetrahydrofolate, *10FTHF* 10-formyltetrahydrofolate, *MET* methionine, *SAM* S-adenosylmethionine, *SAH* S-adenosylhomocysteine, *HCY* homocysteine

**Table 2.1** *The stoichiometric matrix for the folate and methionine cycles metabolism.* Only part of the values of the matrix is shown

|         | 1     | 2    | 3     | 4     | 5     | 6       | 7      | 8   | 9       | 10    | ... | 20    |
|---------|-------|------|-------|-------|-------|---------|--------|-----|---------|-------|-----|-------|
|         | Vbhmt | Vcbs | Vdnmt | Vgnmt | Vmati | Vmatiii | Vmthfr | Vne | Vaicart | Vdhfr | ... | Metin |
| m5mthf  | 0     | 0    | 0     | 0     | 0     | 0       | 1      | 0   | 0       | 0     | ... | 0     |
| thf     | 0     | 0    | 0     | 0     | 0     | 0       | 0      | -1  | 1       | 1     | ... | 0     |
| dhf     | 0     | 0    | 0     | 0     | 0     | 0       | 0      | 0   | 0       | -1    | ... | 0     |
| ch2f    | 0     | 0    | 0     | 0     | 0     | 0       | -1     | 1   | 0       | 0     | ... | 0     |
| chf     | 0     | 0    | 0     | 0     | 0     | 0       | 0      | 0   | 0       | 0     | ... | 0     |
| m10fthf | 0     | 0    | 0     | 0     | 0     | 0       | 1      | 0   | -1      | 0     | ... | 0     |
| met     | 1     | 0    | 0     | 0     | -1    | -1      | 0      | 0   | 0       | 0     | ... | 1     |
| sam     | 0     | 0    | -1    | -1    | 1     | 1       | 0      | 0   | 0       | 0     | ... | 0     |
| sah     | 0     | 0    | 1     | 1     | 0     | 0       | 0      | 0   | 0       | 0     | ... | 0     |
| hcy     | -1    | -1   | 0     | 0     | 0     | 0       | 0      | 0   | 0       | 0     | ... | 0     |

This is because this flux was assumed to be constant in the network. As will be seen later, however (see Sect. 2.5.3), even though metin is a constant, varying this value can result in changes throughout the gradient matrix.

The rank, as well as the size of the row and column spaces, of the stoichiometric matrix is 9, and the rank of the gradient matrix is 10. Given the dimensions and rank of each of these matrices, the size of the null spaces can be calculated (Table 2.3). The single dimension in the left null space of **S** reflects conservation of folate within the network. Since folate is never directly synthesized or degraded in this network, it appears in the left null space.

As discussed in Sect. 2.4, the entries in the stoichiometric matrix are integers, whereas those in the gradient matrix are real valued. The condition number for the

**Table 2.2** The gradient matrix for the network with a methionine input flux of 200  $\mu$ M/h. Only part of the values of the matrix is shown. Note that since the methionine input flux is a constant, all of its entries in the gradient matrix are 0

|     |         | m5mthf | thf   | dhf   | ch2f   | chf | m10fthf | met    | sam    | sah    | hcy   |
|-----|---------|--------|-------|-------|--------|-----|---------|--------|--------|--------|-------|
| 1   | Vbhmt   | 0      | 0     | 0     | 0      | 0   | 0       | 0      | −0.090 | −0.090 | 19.31 |
| 2   | Vcbs    | 0      | 0     | 0     | 0      | 0   | 0       | 0      | 0.077  | 0.077  | 104.2 |
| 3   | Vdnmt   | 0      | 0     | 0     | 0      | 0   | 0       | 0      | 0.228  | −1.257 | 0     |
| 4   | Vgnmt   | −67.46 | 0     | 0     | 0      | 0   | 0       | 0      | 0.294  | −3.772 | 0     |
| 5   | Vmati   | 0      | 0     | 0     | 0      | 0   | 0       | 0.746  | −0.230 | 0      | 0     |
| 6   | Vmatiii | 0      | 0     | 0     | 0      | 0   | 0       | 1.805  | 0.885  | 0      | 0     |
| 7   | Vmthfr  | 0      | 0     | 35.49 | 0      | 0   | 0       | −0.231 | −0.231 | 0      | 0     |
| 8   | Vne     | 0      | 150.0 | 0     | −23.20 | 0   | 0       | 0      | 0      | 0      | 0     |
| 9   | Vaicart | 0      | 0     | 0     | 0      | 0   | 47.80   | 0      | 0      | 0      | 0     |
| 10  | Vdhfr   | 0      | 0     | 8847  | 0      | 0   | 0       | 0      | 0      | 0      | 0     |
| ... | ...     | ...    | ...   | ...   | ...    | ... | ...     | ...    | ...    | ...    | ...   |
| 20  | Metin   | 0      | 0     | 0     | 0      | 0   | 0       | 0      | 0      | 0      | 0     |

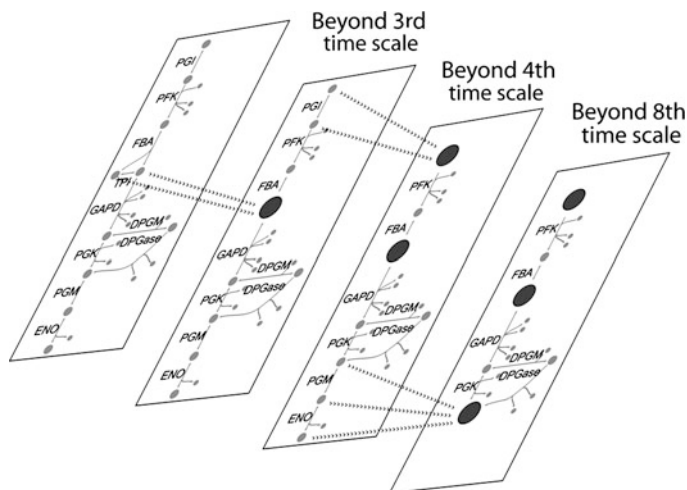
**Table 2.3** The dimensions of the stoichiometric and gradient matrices and the sizes of their right and left null subspaces

|   | Rows | Columns | Left Null Space | Right Null Space |
|---|------|---------|-----------------|------------------|
| S | 10   | 20      | 1               | 11               |
| G | 20   | 10      | 10              | 0                |

system described by theses matrices is approximately  $7.8 \times 10^4$ . This is a relatively large number and reflects the fact that there is a wide range of concentrations for different metabolites in the network and that some of the biochemical interactions in the network occur much quicker (or slower) with respect to other reactions within the network.

2.5.2 Hierarchical Pooling of Metabolites

As discussed in Sect. 2.2.2, a characteristic feature of metabolism, particularly in higher order organisms, is aggregate pool formation of metabolites when one moves from very fast to very slow time scales (Palsson and Lightfoot 1984). This concept is illustrated in Fig. 2.4 for the glycolytic pathway. For this example pooling between chemical isomers occurs on the earliest time scales (these time scales are all faster than milliseconds). There are chemical as well as physiological relevance to pooling of metabolites, and this process occurs in an organized, hierarchical manner. One challenge in biology is to understand this process, because being able to pool metabolites of interest on a particular time scale enables modularization and simplification of an otherwise complex set of interactions. Furthermore, once the network is modularized, it may be possible to identify metabolites (or summed grouped

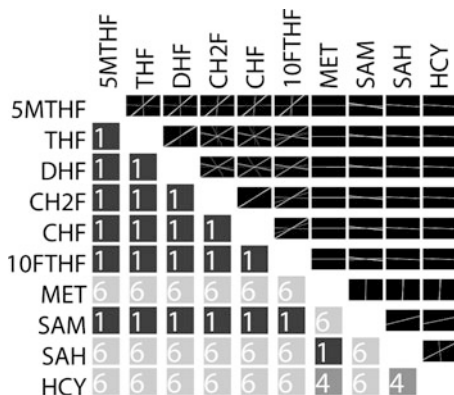


**Fig. 2.4** *Beginning from the fastest time scale and moving forward.* Beginning from the fastest time scale and moving forward, components that move together on subsequent time scales are lumped into an aggregate pool variable. A hierarchical reduction of this network is shown in Fig. 2.6

of metabolites) that reflect different “functional states.” This may help reduce the number of experimental measurements required to characterize the function of a particular network.

Aggregate pools of metabolites can be identified using different approaches. If there are a small number of possible perturbations of interest, then simulation-driven methods can be used (Kauffman et al. 2002). This approach is limited, however, if one wants to characterize all possible responses of the system. An alternative is to adopt an analytical approach through the analysis of the Jacobian around a particular steady state (Jamshidi and Palsson 2008b). The network can be dynamically decoupled, and then any correlations between metabolites (or fluxes) can be assessed on every single one of the independent time scales. Through calculation of correlations between all of the metabolites (or fluxes) on progressive time scales, removing the time scales one by one (beginning with the fastest), and recalculating correlations between the components, one can identify the pools that form. This procedure (see Jamshidi and Palsson (2008b)) was carried out for the folate network with a methionine input flux of 200  $\mu\text{M}/\text{min}$  and are visually depicted in Fig. 2.5.

There were seven independent time scales for the network under these conditions, and there is clear separation of the folate carrier branch from the methionine cycle, although SAM immediately pools with the folate metabolites. Analysis of the kinetics in the context of the stoichiometric matrix identified that the pooling of SAM with the folate cycle was not stoichiometric determined (i.e., there are no reactions that directly involve metabolites from the folate cycle and SAM), but these were kinetically driven events.



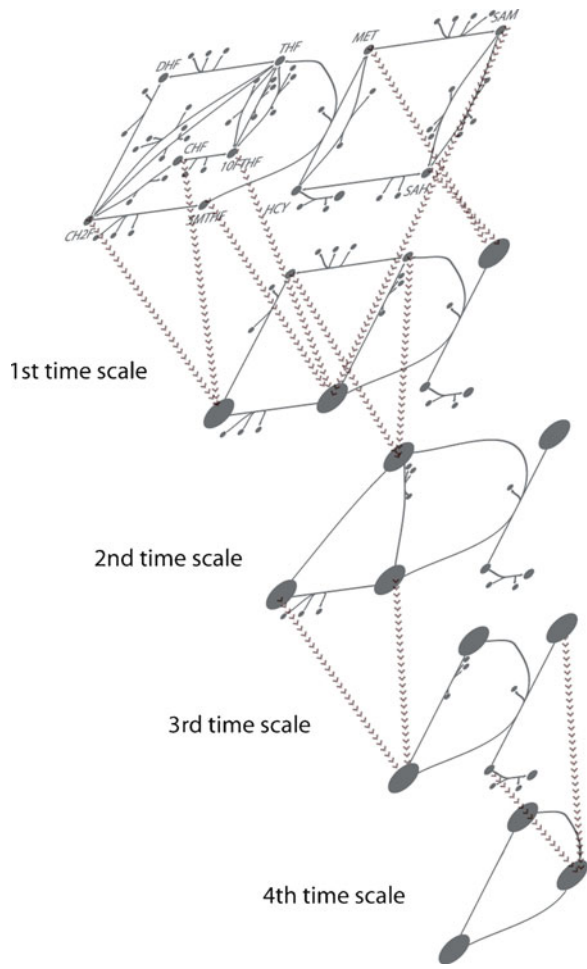
**Fig. 2.5** *Pooling among metabolites on progressive time scales.* Methionine input flux at  $200\ \mu\text{M}/\text{h}$ . Time scale hierarchy of metabolic pool formation in the human red blood cell. The *lower left triangle* indicates the modes after which pooling occurs between the corresponding metabolites (one being the fastest time scale). The *upper right triangle* are plots of the slopes between the two metabolites for the remaining time scales after pool formation (the origin is always included in these approximations of the slopes), color coded according to the time scale at which pooling occurs. A correlation cutoff of 0.9 was used for the pooling criteria determination

### 2.5.3 Environmental Perturbations

A benefit of building a model *in silico* is the ability to carry out various perturbations and to observe the changes that occur. The methionine input flux is described by a zero-order rate expression, and as noted in Table 2.2, all of its entries in the gradient matrix are 0. However, changes in the methionine input will cause the system to shift from one steady state to another. This change may result in altered network dynamics. The methionine input flux was considered at halved as well as doubled rates. A cursory glance at the numerical entries shows that many of the values are significantly different under the different conditions. This implies that different homeostatic states have different dynamic properties and quantitatively different systemic – in response to perturbations.

One can immediately see differences in the entries of the gradient matrix, as well as the **K** and **M** matrices for these different conditions. The network-wide changes are more easily highlighted in the tiled pooling arrays of the networks for methionine input fluxes of  $400$  and  $50\ \mu\text{M}/\text{h}$ , as shown in Figs. 2.7 and 2.8. At these alternate methionine input flux states, the pooling among metabolites has completely changed. Most notably pooling within the folate cycle occurs much later at the lower methionine input flux rate. At the much higher methionine input flux rate, we see that there are effectively two time scales in which pooling occurs, the first time scale ( $\sim 0.5\ \text{ms}$ ) and the seventh time scale ( $\sim 45\ \text{s}$ ).

These results highlight not only the importance of environmental conditions in the analysis of dynamics in metabolic networks but also the potential for different dynamic properties at different steady states in networks.

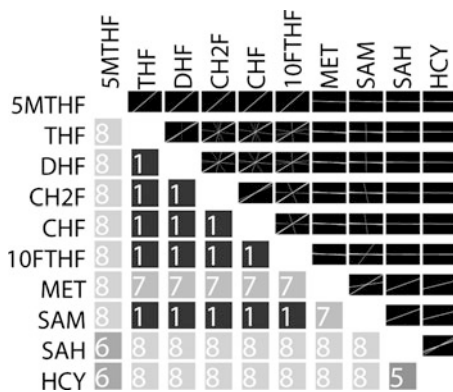


**Fig. 2.6** Hierarchical reduction of the network in Fig. 2.3. Progressive pooling of metabolites in folate and methionine cycles was determined according to Fig. 2.5

**Fig. 2.7** Pooling among metabolites on progressive time scales. Methionine input flux at 50  $\mu$ M/h

|        | 5MTHF | THF | DHF | CH2F | CHF | 10FTHF | MET | SAM | SAH | HCY |
|--------|-------|-----|-----|------|-----|--------|-----|-----|-----|-----|
| 5MTHF  |       |     |     |      |     |        |     |     |     |     |
| THF    | 5     |     |     |      |     |        |     |     |     |     |
| DHF    | 5     | 2   |     |      |     |        |     |     |     |     |
| CH2F   | 5     | 5   | 5   |      |     |        |     |     |     |     |
| CHF    | 5     | 5   | 5   | 1    |     |        |     |     |     |     |
| 10FTHF | 1     | 5   | 5   | 3    | 3   |        |     |     |     |     |
| MET    | 6     | 6   | 6   | 6    | 6   |        |     |     |     |     |
| SAM    | 1     | 5   | 5   | 5    | 5   | 1      | 6   |     |     |     |
| SAH    | 6     | 6   | 6   | 6    | 6   | 6      | 1   | 6   |     |     |
| HCY    | 6     | 6   | 6   | 6    | 6   | 6      | 4   | 6   | 4   |     |

**Fig. 2.8** Pooling among metabolites on progressive time scales. Methionine input flux at 400  $\mu\text{M/h}$



## 2.6 Conclusions

To date there has not been a successful, generalized strategy to build genome-scale kinetic models. This has been principally due to the large number kinetic parameters required to define the system which is further confounded by the fact that in vitro measurements of kinetic constants are often not representative of their numerical values in vivo. These challenges have lead to the infeasibility of achieving cell scale models using such approaches. Identification of the key structural and dynamic properties of networks and the inherent relationships between fluxes and concentrations will help to achieve dynamic descriptions of genome-scale models. Here, we showed how the dynamics of a biochemical reaction network can be described by dual Jacobian matrices, which is enabled by recognition of the fact that dynamic interactions are constrained by network topology. Fluxes and concentrations are dual variables in biochemical reaction networks, but they are related via changes in fluxes and concentrations. These relationships are described by the gradient matrix. The ability to convert from one set of variables into another is not just of mathematical interest, but highlights the underlying roots of the relationship between fluxes and concentrations. Ultimately the characterization of biological systems is to understand how the system responds to perturbations. To date, dynamic descriptions of networks have been confined to the column space; however, the relationships described here allow one to describe the network in terms of column space or row space variables. This is of particular interest in biological networks, as the perturbation variables are generally concentration variables in the column and left null spaces, whereas the response variables are the fluxes, in the row and right null spaces. Thus, concentration variables perturb a network, and the flux variables respond to the perturbation and tie the network together. These are complementary variables that are tied together in the network by the stoichiometric and gradient matrices.

A key motivation in being able to build larger models is to then analyze, understand, and hopefully simplify these networks. Herein, we focused on some of the approaches for simplification of network dynamics in the context of dynamic

hierarchies. A goal in these efforts is to window in on a time scale of interest and to determine the simplified, pooled structure of a network. This effectively filters out processes that occur too slowly or too quickly to be of interest and may highlight grouped metabolites that can be used as surrogates for network functional states, such as the redox state or energy charges of a cell.

### 2.6.1 Future Directions: Constructing Genome-Scale Models

Previously when models of biochemical reactions and networks have been constructed, it has been through the statement of assumptions such as quasi-equilibrium and quasi-steady state, followed by incorporation of data into the models and curve fitting parameters; thus the statements of assumptions are in effect “preprocessing” the model. The description and decomposition of models described here are carried out from a different perspective. The mechanistic, bilinear interactions are represented in the stoichiometric matrix, the various high-throughput data types (nucleic acid, protein, and small metabolite concentrations) are incorporated into the model, and then the decisions are made for assumptions. These assumptions can be varied and adjusted depending on the question of interest and the time scale(s) of interest. Thus, this is a mechanistic, data-driven approach, in which assumptions are a “postprocessing” step of model construction. This approach recognizes and appreciates the stoichiometric and gradient matrices as the key matrices in building large-scale networks. There has been some progress in this area for outlining the approaches to build kinetic models. We have recently developed an approach that is practical, feasible, and successful in test cases to date (Jamshidi and Palsson 2010). As progress continues to be made in the “-omics” field, particularly metabolomics, we anticipate the development of genome-scale kinetic models in the near future.

## Appendix: Details About Matrices

### Forming the Gradient Matrix

Dynamic analysis of complex systems is normally carried out with the linearization of the right-hand side of (2.1). Noting that  $\mathbf{S}$  is a matrix with constant coefficients, linearization of (2.1) comes down to the Taylor series expansion of reaction rates  $\mathbf{v}(\mathbf{x})$ :

$$\mathbf{v}(\mathbf{x}) = \mathbf{v}(\mathbf{x}_0) + \left. \frac{d\mathbf{v}}{d\mathbf{x}} \right|_{\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} \left. \frac{d^2\mathbf{v}}{d\mathbf{x}^2} \right|_{\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0)^2 + \dots \quad (2.6)$$

Neglecting all second order and higher terms yields,

$$\mathbf{v}(\mathbf{x}) \approx \mathbf{v}(\mathbf{x}_0) + \left. \frac{d\mathbf{v}}{d\mathbf{x}} \right|_{\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0) \quad (2.7)$$

When the reference state,  $\mathbf{x}_0$ , is specified as a steady state for the system, then by definition,

$$\mathbf{S} \cdot \mathbf{v}(\mathbf{x}_0) = \mathbf{0} \quad (2.8)$$

so that the linearized form of (2.1) is,

$$\frac{d(\mathbf{x} - \mathbf{x}_0)}{dt} = \mathbf{S} \cdot \left. \frac{d\mathbf{v}}{d\mathbf{x}} \right|_{\mathbf{x}_0} \cdot (\mathbf{x} - \mathbf{x}_0) \quad (2.9)$$

So quite naturally one can define the gradient matrix,  $\mathbf{G}$ ,

$$\mathbf{G} = \frac{d\mathbf{v}}{d\mathbf{x}} \quad (2.10)$$

We note that this is not an arbitrary or a definition of mathematical convenience, but simply the result of linearization of fluxes around a specified reference point. We further note that the gradient matrix is equal to the nonlogarithmic form of the elasticity matrix in metabolic control analysis (Hatzimanikatis and Bailey 1996).

The stoichiometric matrix has been investigated in detail in the literature (Palsson 2006). Since the gradient matrix has only recently been recognized (Jamshidi and Palsson 2008a), time will be spent highlighting and contrasting its key features with the stoichiometric matrix.

### *The Jacobian Matrix for Concentrations*

Specifying the reference point in (2.7),  $\mathbf{x}_0$ , to be a steady state for the system, (2.1) becomes:

$$\frac{d\mathbf{x}'}{dt} = \mathbf{S} \cdot \mathbf{G} \cdot \mathbf{x}' \quad (2.11)$$

in which  $\mathbf{x}'$  is the deviation variable,  $(\mathbf{x} - \mathbf{x}_0)$ .  $\mathbf{J}^x = \mathbf{S} \cdot \mathbf{G}$  is the Jacobian for the system of equations describing the concentration variables. Note that this factorization separates the chemistry that specifies network topology (through  $\mathbf{S}$ ), and the kinetics and thermodynamics that give the driving forces and their time scale of action (residing in  $\mathbf{G}$ ). These two effects can be effectively separated by scaling the rows of  $\mathbf{G}$  to unity as (Jamshidi and Palsson 2008a):

$$\mathbf{G} = \mathbf{K}^v \cdot \mathbf{M}^v \quad (2.12)$$

where the rows in  $\mathbf{M}^v$  represent the direction of the driving forces (the thermodynamics) in the row space.  $\mathbf{M}^v$  is a row-normalized gradient matrix, and each row corresponds to a reaction. The matrix  $\mathbf{K}^v$  is diagonal. Its elements represent the time scales on which the thermodynamic force of a reaction acts, of the kinetics. In this formulation, the rows of the gradient matrix are drivers and the columns of the stoichiometric matrix define the directions of motion.

## The Jacobian Matrix for Fluxes

The concentrations and fluxes are two sets of variables that characterize the dynamic state of a network. Either can in principle be used as the set of independent variables and the other computed as set of dependent variables. Stoichiometric matrices for biochemical networks are, however, normally rectangular with  $m < n$ , and rank,  $r < m$ .  $\mathbf{S}$  is thus not invertible and (2.1) cannot be directly converted into a system of dynamic equations in terms of fluxes.

The gradient matrix enables the change of the system of equations from the concentration variables to a system of equations in terms of flux variables. Defining the flux deviation variable,  $\mathbf{v}' = \mathbf{G} \cdot \mathbf{x}'$ , and premultiplying (2.9) by the gradient matrix yields:

$$\frac{d\mathbf{v}'}{dt} = \mathbf{G} \cdot \mathbf{S} \cdot \mathbf{v}' \quad (2.13)$$

Thus the Jacobian matrix is  $\mathbf{J}^v = \mathbf{G} \cdot \mathbf{S}$ , when treating the fluxes as the independent variables. In a similar way as above, we can scale every column in  $\mathbf{J}^v$  and factor the gradient matrix as:

$$\mathbf{G} = \mathbf{M}^x \cdot \mathbf{K}^x \quad (2.14)$$

yielding  $\mathbf{J}^v = \mathbf{M}^x \cdot \mathbf{K}^x \cdot \mathbf{S}$ . Here,  $\mathbf{M}^x$  has the columns of  $\mathbf{G}$  normalized to unity, and the diagonal matrix  $\mathbf{K}^x$  contains the length of these columns, which correspond to compounds. Note that the elements of  $\mathbf{M}^x$  represent the kinetic potential of compounds.

$\mathbf{J}^v$  is thus reassembled compound by compound, whereas  $\mathbf{J}^x$  was assembled reaction by reaction. In this formulation, drivers (the rows of  $\mathbf{S}$ ) are the sums of the fluxes in and out of a node multiplied by the kinetic potential of the compound. The directions of motions are given by the columns of  $\mathbf{M}^x$ , and the elements in the diagonal matrix  $\mathbf{K}^x$  determine the weights or influence of the motions. The direction of a column in  $\mathbf{M}^x$  designates the kinetically balanced outflow of a compound from a node, if the concentration of the compound in that node is perturbed from steady state.

## References

- I. Famili and B. Ø. Palsson. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys J*, 85:16–26, 2003
- V. Hatzimanikatis and J. Bailey. MCA has more to say. *J Theor Biol*, 182:233–242, 1996
- H. Heinrich and S. Schuster. *The regulation of cellular systems*. Springer, Heidelberg, 1996
- R. Heinrich, S. M. Rapoport, and T. A. Rapoport. Metabolic regulation and mathematical models. *Prog Biophys Mol Biol*, 32:1–82, 1977
- N. Jamshidi and B. Ø. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, 4:171, 2008a

- N. Jamshidi and B. Ø. Palsson. Top-down analysis of temporal hierarchy in biochemical reaction networks. *PLoS Comput Biol*, 4:e1000177, 2008b
- N. Jamshidi and B. Ø. Palsson. Mass action stoichiometric simulation models: Incorporating kinetics and regulation into stoichiometric models. *Biophys J*, 98(2):175–185, 2010
- K. J. Kauffman, J. D. Pajewski, N. Jamshidi, B. Ø. Palsson, and J. S. Edwards. Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophys J*, 83:646–662, 2002
- E. Mayr. Cause and effect in biology. *Science*, 124:1501–1506, 1961
- M. S. Okino and M. L. Mavrouniotis. Simplification of mathematical models of chemical reaction systems. *Chem Rev*, 98:391–408, 1998
- B. Ø. Palsson. *Systems biology: Determining the capabilities of reconstructed networks*. Cambridge University Press, Cambridge, 2006
- B. Ø. Palsson and E. N. Lightfoot. Mathematical modelling of dynamics and control in metabolic networks. I. On Michaelis-Menten kinetics. *J Theor Biol*, 111:273–302, 1984
- M. C. Reed, H. F. Nijhout, M. L. Neuhouser, J. F. Gregory, B. Shane, S. J. James, A. Boynton, and C. M. Ulrich. A mathematical model gives insights into nutritional and genetic aspects of folate-mediated one-carbon metabolism. *J Nutr*, 136:2653–2661, 2006
- J. Reich and E. Selkov. *Energy metabolism of the cell: A theoretical treatise*. Academic, London, 1981
- I. Segel. *Enzyme kinetics*. Wiley, New York, 1975
- G. Strang. *Linear algebra and its applications*. Harcourt Brace Jovanovich, San Diego, 1988

# Chapter 3

## A Deterministic, Mathematical Model for Hormonal Control of the Menstrual Cycle

R. Drew Pasteur and James F. Selgrade

### 3.1 Introduction and Biological Background

The reproductive endocrine system is one of the most complex systems in the human body. In women, there are three sources of hormone production: the hypothalamus, the pituitary gland, and the ovaries. The hormones produced in these three locations jointly regulate the processes occurring at all three sites and control the processes surrounding ovulation and menstruation, primarily by feedback from the ovaries to the pituitary (Speroff et al. 1999). Figure 3.1 summarizes the interactions of these feedback loops.

Pharmaceutical use of external hormones is common across the life cycle, most notably for contraception or treatment of abnormal menstrual cycles in younger women, and later to suppress the undesirable effects of menopause. Hormone-like substances, particularly xenoestrogens, can also be unintentionally ingested, via food (Davis et al. 1993) and drinking water (Rudel et al. 1998). Because breast cancer risk may be related to total lifetime exposure to bioavailable estrogens, there is rising concern over the long-term dangers of estrogen exposure, regardless of the source (Davis et al. 1993). A mathematical model can be used to illustrate hormone levels which prevent ovulation; this is discussed in Sect. 3.5.

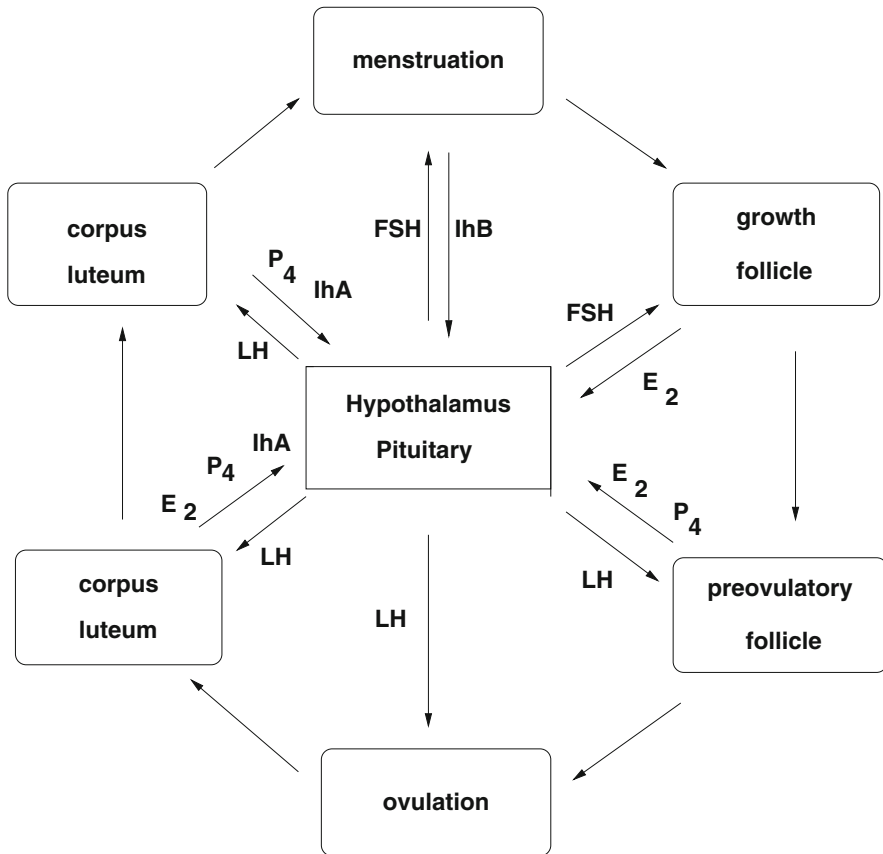
While a deterministic mathematical model cannot fully predict the results of external influences on the reproductive endocrine system, it might identify a potential course of action to be considered in future clinical research studies. Recent mathematical models, such as Clark et al. (2003) and Reinecke and Deuflhard (2007), are substantially more complex than their predecessors, taking full advantage of modern high-speed computers.

The availability of separate bioassays for two forms of the ovarian hormone inhibin is relatively new, with full-cycle data first appearing in Groome et al. (1996). Later studies involving inhibin B, such as the one reported in Welt et al. (1999),

---

R.D. Pasteur (✉)

Department of Mathematics and Computer Science, The College of Wooster,  
1189 Beall Ave., Wooster, OH 44691, USA  
e-mail: [rpasteur@wooster.edu](mailto:rpasteur@wooster.edu)



**Fig. 3.1** *Phases of the menstrual cycle.* The phases displayed indicate interactions among gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH), and ovarian hormones estradiol (E<sub>2</sub>), progesterone (P<sub>4</sub>), inhibin A (IhA), and inhibin B (IhB). A similar figure was published first in *Fields Inst Comm* (Selgrade and Schlosser 1999) by the AMS

had larger numbers of subjects and considered women from multiple age groups. In addition to inhibin B (IhB), the ovaries also produce inhibin A (IhA), estradiol (E<sub>2</sub>, a form of estrogen), progesterone (P<sub>4</sub>), and other hormones. Ovarian hormone production gradually declines as women approach menopause. For the inhibins, this decrease starts early, around age 35 (Welt et al. 1999). A more comprehensive mathematical model has the potential to give insight into hormonal issues in perimenopausal women, not just the younger women represented by the current model (Clark et al. 2003).

Both E<sub>2</sub> and P<sub>4</sub> are steroid hormones, produced from acetate and cholesterol via the same chemical pathway as testosterone. The hypothalamus secretes gonadotropin-releasing hormone (GnRH), which leads to the production of follicle-stimulating hormone (FSH) and luteinizing hormone (LH), both of which are gonadotropins, by the pituitary (Speroff et al. 1999).

In the absence of exogenous hormones, the bloodstream hormone levels of women of childbearing age are not at equilibrium, but rather change in a predictable, periodic manner. Together these dynamic interactions control the menstrual cycle. While the average human menstrual cycle for a healthy, fertile woman is 28 days, most women do not have a 28-day cycle, and cycles several days longer or shorter can be considered normal (Speroff et al. 1999; Vollman 1977; Treloar et al. 1967). Adolescents in their early teens, as well as women approaching menopause, tend to have longer, irregular cycles, during their first and last few menstrual years, respectively. Ovulation, which is necessary for fertility, may or may not occur during these long, irregular cycles (Yen 1999; Vollman 1977; Treloar et al. 1967). The menstrual cycle is stopped temporarily by oral contraceptives or pregnancy and is permanently ended naturally at menopause, or sooner via a hysterectomy.

The menstrual cycle can be tracked by noting the dates of the start of menstruation, vaginal bleeding typically lasting several days. The first day of menstruation is conventionally considered the first day of a new menstrual cycle. Ovulation occurs roughly halfway through the cycle, splitting it into two phases; the first half is called the follicular phase, and the latter half the luteal phase (Fig. 3.1).

At ovulation, a mature follicle ruptures, releasing an egg for possible fertilization. The developmental processes leading up to that point last for roughly 85 days (Oktay et al. 1998); so at any time, there are multiple active follicles at various stages of maturity, regardless of the phase of the menstrual cycle (Baird 1984). At puberty, each ovary contains roughly 250,000 immature egg cells called oocytes, but only a few hundred of these will ever reach full follicular maturity and rupture at ovulation (Hadley 1992). The others will at some point undergo atresia, a degenerative process; eventually, menopause occurs when no active follicles remain (Speroff et al. 1999). In a healthy woman not taking hormonal contraceptives, typically only one follicle develops to the point of ovulation during each menstrual cycle. However, the incidence of fraternal twins, due to multiple fertilized eggs, demonstrates the possibility of more than one follicle completing the developmental process during a given cycle.

Shortly before menstruation, roughly 2–3 weeks before the next expected ovulation, multiple follicles in each ovary begin an advanced state of development, due to slowly rising levels of FSH. Each such follicle has the potential of rupturing in ovulation during the upcoming cycle (Pache et al. 1990). During the week after menstruation, one follicle begins to grow exponentially, while the others begin to break down through atresia (Baird 1984). The physiological processes by which one follicle becomes dominant are not well understood (Zeleznik and Pohl 2006).

FSH levels are elevated throughout the follicular phase, as this is a necessary condition for the advanced follicular development taking place in the maturing primary follicle. During the latter week of the follicular phase, this follicle releases E2 in large quantities, to the point that there is an exponential rise in bloodstream levels of E2 (Baird 1984). The high E2 levels reduce the release of FSH by the pituitary, which eventually causes a sufficient drop in ovarian FSH levels to accelerate the demise of less-developed cohort follicles (Speroff et al. 1999). The primary follicle is unaffected due to internal storage of FSH (Baird 1984).

Throughout the follicular phase, a large amount of LH is stored in reserve. It is then emptied into the bloodstream over about 2 days. The resulting high-amplitude spike in circulating LH levels, peaking at eight times the baseline level, on average (Welt et al. 1999), ultimately leads to ovulation. Simultaneously, there is a significant release of FSH; while less dramatic, it also has physiologic implications related to ovulation. According to the biological literature, the LH surge is triggered by sustained high concentrations of E2, above a threshold level for at least 36–48 h (Speroff et al. 1999; Young and Jaffe 1976). However, at subthreshold levels, increased E2 causes a decrease in LH release; so LH does not vary monotonically with E2. Our model included dual control of LH by E2, with negative feedback on LH release, but positive feedback on LH synthesis at levels above a threshold (Liu and Yen 1983). As E2 increases sharply during the late follicular phase, we modeled the resulting promotion of LH synthesis as being primarily responsible for the LH surge.

Concentrations of circulating P4 are low throughout the follicular phase, but rise slowly just before mid-cycle, and likely have a role in the timing of the LH surge (Speroff et al. 1999). Circulating levels of one more hormone, IhB, also peak at mid-cycle, possibly due to a substantial release from the ruptured follicle at ovulation, but this is unclear (Speroff et al. 1999; Muttukrishna et al. 2000).

Following ovulation, the luteal phase lasts roughly 2 weeks for a typical healthy woman. During this phase, the ruptured primary follicle, now called the corpus luteum, releases large amounts of P4, preparing the body for a potential pregnancy (Baird 1984). Within a few days after ovulation, IhB levels drop substantially, and IhB has no significant role in regulating FSH during this time. However, IhA concentrations, already rising since the mid-follicular phase, peak during the mid-luteal phase. There is an extended, low-amplitude peak of E2 during the luteal phase; together, E2 and IhA suppress the production of FSH during the latter half of each cycle (Welt et al. 1999; Yen 1999).

### 3.2 The Pituitary and Ovarian Models

In modeling the reproductive endocrine system in women, we began with separate models of the relevant processes occurring in the pituitary gland and the ovaries. Pituitary hormone production is pulsatile in nature, in response to pulsatile stimulation by GnRH from the hypothalamus (Speroff et al. 1999). However, following the models presented by Schlosser and Selgrade (2000), we lumped together effects of the hypothalamus and pituitary gland, and smoothed out the pulsatile effects by considering average synthesis, lengthening the time scale. As discussed previously, when LH and FSH are produced in the anterior pituitary, there is not an immediate full release of these hormones into the bloodstream upon synthesis. Hence, we provided for a pituitary reserve called a releasable pool, which is a biologically reasonable assumption based on Speroff et al. (1999).

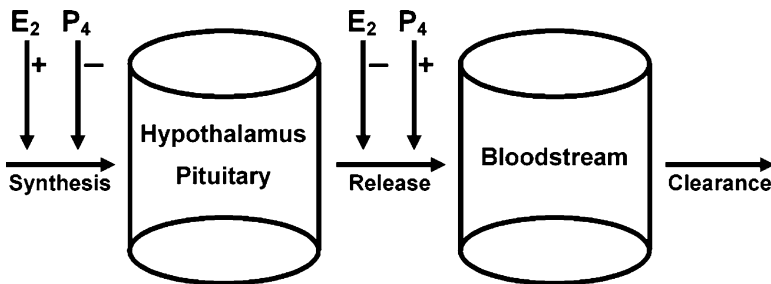
To model the pituitary hormones LH and FSH, we tracked for each the amount of hormone in the releasable pool, as well as the concentration of hormone in the bloodstream. This required quantifying the rates of three biological processes for each hormone: synthesis, release, and clearance. Because synthesis and release rates depend on the circulating concentrations of the four ovarian hormones, we constructed periodic functions of time to represent the concentrations of E2, P4, IhA, and IhB, fitting the data from [Welt et al. \(1999\)](#). These functions served as inputs into the systems of differential equations modeling LH and FSH.

Because there are observed delays between various peaks and valleys in the graphs of ovarian hormones and the expected corresponding peaks and valleys of graphs of LH and FSH, in [Welt et al. \(1999\)](#) as well as in [McLachlan et al. \(1990\)](#), we used constant time delays to represent such effects.

Following [Schlosser and Selgrade \(2000\)](#), we modeled LH synthesis, release, and clearance with the system of differential equations [(3.1) and (3.2)]. The state variables  $RPLH$  and  $LH$  represent the amount of LH in the releasable pool and the LH concentration in the blood, respectively; the parameter  $v$  denotes blood volume. Figure 3.2 is a schematic diagram for the hypothalamus/pituitary system and the influences on it by the ovaries.

$$\begin{aligned} \frac{d}{dt} RPLH &= \text{LH Synthesis}(E2, P4) - \text{LH Release}(E2, P4, RPLH) \\ \frac{d}{dt} LH &= \frac{1}{v} \text{LH Release}(E2, P4, RPLH) - \text{LH Clearance}(LH) \end{aligned} \quad (3.1)$$

As discussed in the previous section, E2 and P4 have mutually antagonistic effects on the synthesis and release of LH. Also noteworthy in (3.2) is the numerator of the LH synthesis term, which includes a rational function called a Hill function. Physiological research suggests that a sustained high concentration of E2 is needed in the bloodstream for at least 36–48 h, to trigger the LH surge, which in turn causes ovulation to occur ([Speroff et al. 1999](#); [Young and Jaffe 1976](#)). It is unclear whether



**Fig. 3.2** Schematic diagram for the LH system given by (3.1) and (3.2). Arrows represent regulatory feedback of a given hormone on a particular physiologic process. Positive and negative signs indicate promotion and inhibition, respectively, of that process

this E2 effect controls the synthesis or the release of LH, but our model includes it in the synthesis term. The parameter  $Km_{LH}$  is a threshold value for E2 at which half-maximal LH synthesis occurs. As the exponent  $a$  increases, the gradient for LH synthesis becomes steep as the E2 concentration is near this threshold level. In our model, for high values of  $a$ , relatively little LH synthesis (above the baseline level represented by  $v_{0,LH}$ ) occurred for E2 levels below  $Km_{LH}$ , but production increased sharply once the concentration of E2 exceeded this value.

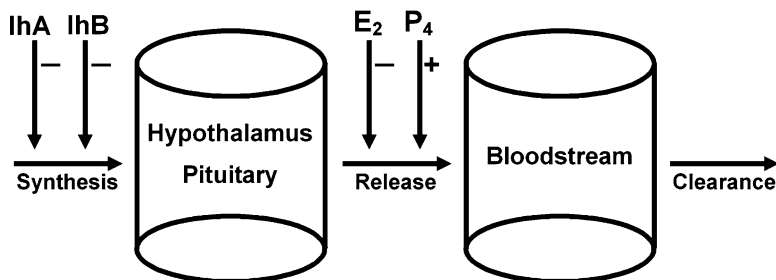
$$\begin{aligned}
 \text{LH Synthesis(E2, P4)} &= \frac{v_{0,LH} + v_{1,LH} \cdot \frac{(E2(t-d_E)/Km_{LH})^a}{1 + (E2(t-d_E)/Km_{LH})^a}}{1 + \frac{P4(t-d_P)}{Ki_{LH,P}}} \\
 \text{LH Release(E2, P4, RPLH)} &= k_{LH} \cdot \frac{1 + c_{LH,P} \cdot P4(t)}{1 + c_{LH,E} \cdot E2(t)} \cdot \text{RPLH}(t) \\
 \text{LH Clearance(LH)} &= r_{LH} \cdot \text{LH}(t)
 \end{aligned} \tag{3.2}$$

Of the parameters in (3.1) and (3.2), only two were estimated from the physiological literature. For the clearance rate of LH from the bloodstream, we used  $r_{LH} \approx 14 \text{ day}^{-1}$ , based on Kohler et al. (1968), and for the volume of circulating blood, we used  $v \approx 2.5 \text{ L}$ . The delays associated with the effects of E2 and P4 on LH synthesis,  $d_E$  and  $d_P$ , were expected to be on the scale of 0–2 days, and a beginning estimate of the E2 synthesis threshold value  $Km_{LH}$  was taken from the bloodstream concentrations of E2 given in Welt et al. (1999), but none of the other parameters were known.

The model for FSH in the pituitary, given by (3.3), is similar in form to that of (3.1), but without a Hill function, because there is no physiological evidence that either FSH synthesis or release is as sensitive as LH to small changes in concentration of any ovarian hormone. Both IhA and IhB inhibit synthesis of FSH, and the release of FSH is progressively stifled as concentrations of E2 rise. Just as with LH, high levels of P4 lead to an increase in the relative release rate of FSH. Equations (3.3) and (3.4) model the synthesis, release, and clearance of FSH (Fig. 3.3).

$$\begin{aligned}
 \frac{d}{dt} \text{RPFOSH} &= \text{FSH Synthesis(IhA, IhB)} - \text{FSH Release(E2, P4, RPFOSH)} \\
 \frac{d}{dt} \text{FSH} &= \frac{1}{v} \text{FSH Release(E2, P4, RPFOSH)} - \text{FSH Clearance(FSH)}
 \end{aligned} \tag{3.3}$$

$$\begin{aligned}
 \text{FSH Synthesis(IhA, IhB)} &= \frac{v_{FSH}}{1 + \frac{IhA(t-d_{IhA})}{Ki_{FSH,IhA}} + \frac{IhB(t-d_{IhB})}{Ki_{FSH,IhB}}} \\
 \text{FSH Release(E2, P4, RPFOSH)} &= k_{FSH} \cdot \frac{1 + c_{FSH,P} \cdot P4(t)}{1 + c_{FSH,E} \cdot E2(t)} \cdot \text{RPFOSH}(t) \\
 \text{FSH Clearance(FSH)} &= r_{FSH} \cdot \text{FSH}(t)
 \end{aligned} \tag{3.4}$$



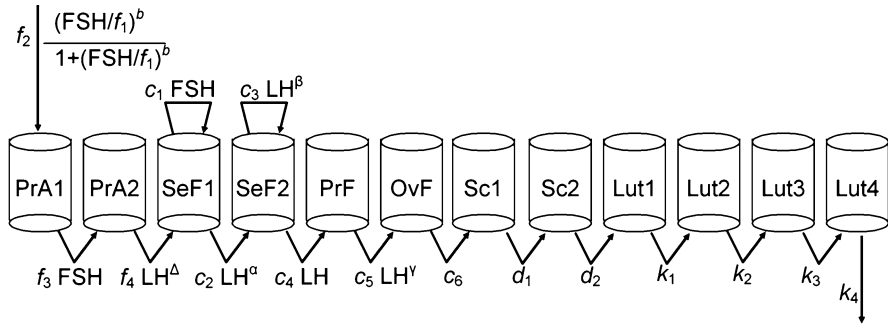
**Fig. 3.3** Schematic diagram for the FSH system given by (3.3) and (3.4). Arrows represent regulatory feedback of a given hormone on a particular physiologic process. Positive and negative signs indicate promotion and inhibition, respectively, of that process

The parameter  $\nu$  in (3.4) is unchanged from the LH system in (3.2). We had a physiological estimate that the clearance rate of FSH from the bloodstream is  $r_{\text{FSH}} \approx 8.21 \text{ day}^{-1}$ , based on Coble et al. (1969). Additionally, we expected that the delays in the effects of the inhibins,  $d_{\text{IhA}}$  and  $d_{\text{IhB}}$ , are on the order of 1–3 days, but all of the other parameters were determined to fit the model to the clinical data best.

In agreement with the previous models of the human menstrual cycle such as Bogumil et al. (1972), as well as Selgrade and Schlosser (1999), we assumed that the clearance rates of the ovarian hormones are much faster than those of the gonadotropins LH and FSH. With this in mind, we used a different approach to model the bloodstream concentrations of the ovarian hormones, considering them to be at quasi-steady state (Keener and Sneyd 2009) because of the difference in time scales. Selgrade and Schlosser (1999) divided the menstrual cycle into nine stages, each representing the capacity of the ovary to produce various hormones, with three stages for the follicular phase, two stages around the time of ovulation, and four stages for the luteal phase. The circulating concentration of each ovarian hormone was then modeled as a linear combination of these nine stages as in (3.6) below, plus a baseline level in some cases. The fast clearance simplifies the model, avoiding the necessity of tracking synthesis, release, and clearance of each hormone.

We used a similar formulation for the ovarian model, with 12 stages instead of nine. Because our model includes an additional hormone IhB, which peaks early in the follicular phase, we needed additional stages that likewise peak very early in the cycle, to represent preantral follicles. Figure 3.4 is a schematic diagram for the 12-stage ovarian model. The compartments represent 12 sequential stages of the menstrual cycle, further breaking down the stages of Fig. 3.1. The mass in a compartment represents the ability of the ovary to produce particular hormones. Arrows from one compartment to another indicate transfer of mass, while arrows from a compartment to itself indicate an exponential growth process within a stage.

Transitions noted with constant coefficients are linear mass transfers; that is, the transfer rate is proportional to the mass of the stage from which the transfer occurs. The transfers in the early stages are dependent on LH and/or FSH, as both pituitary



**Fig. 3.4** Schematic diagram for the 12-stage ovarian model. Arrows represent transfer of mass from 1 stage to another or mass growth within a stage

hormones are important during the follicular phase. Arrows from a stage to itself, in the 3 and 4 stages, represent exponential growth of the primary follicle, during the mid-to-late follicular phase.

Also notable is the Hill function which provides the mass flow for the 1 stage, starting the whole process. The physiological literature suggests that there may be a critical FSH threshold for production of inhibin B during the luteal-follicular transition (Welt et al. 1997). For this reason, we have included another Hill function, and FSH logically is the catalyst. It should be noted that IhB synthesis does not stem directly from the presence of FSH, but rather follows indirectly, as FSH promotes follicular development, eventually leading to IhB production.

Each of the 12 stages corresponds to a differential equation, based on the inflow and outflow of mass at that stage. At any time, the bloodstream concentrations of E2, P4, IhA, and IhB are determined from the 12 stage values by the use of auxiliary (3.6), one for each of the four ovarian hormones. Each auxiliary equation consists of a constant baseline level plus a linear combination of some (but not all) of the 12 stages. Equation (3.5) gives the 12 ordinary differential equations for the ovarian model, with 21 unknown parameters. Equation (3.6) gives the four auxiliary equations which determine the ovarian hormone concentrations in terms of the stages, via 16 additional parameters.

$$\begin{aligned}
 \frac{d}{dt} \text{PrA1} &= f_2 \cdot \frac{(\text{FSH}/f_1)^b}{1 + (\text{FSH}/f_1)^b} - f_3 \cdot \text{FSH} \cdot \text{PrA1} \\
 \frac{d}{dt} \text{PrA2} &= f_3 \cdot \text{FSH} \cdot \text{PrA1} - f_4 \cdot \text{LH}^\delta \cdot \text{PrA2} \\
 \frac{d}{dt} \text{SeF1} &= f_4 \cdot \text{LH}^\delta \cdot \text{PrA2} + (c_1 \cdot \text{FSH} - c_2 \cdot \text{LH}^\alpha) \cdot \text{SeF1} \\
 \frac{d}{dt} \text{SeF2} &= c_2 \cdot \text{LH}^\alpha \cdot \text{SeF1} + (c_3 \cdot \text{LH}^\beta - c_4 \cdot \text{LH}) \cdot \text{SeF2} \\
 \frac{d}{dt} \text{PrF} &= c_4 \cdot \text{LH} \cdot \text{SeF2} - c_5 \cdot \text{LH}^\gamma \cdot \text{PrF}
 \end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}\text{OvF} &= c_5 \cdot \text{LH}^\gamma \cdot \text{PrF} - c_6 \cdot \text{OvF} \\
\frac{d}{dt}\text{Sc1} &= c_6 \cdot \text{OvF} - d_1 \cdot \text{Sc1} \\
\frac{d}{dt}\text{Sc2} &= d_1 \cdot \text{Sc1} - d_2 \cdot \text{Sc2} \\
\frac{d}{dt}\text{Lut1} &= d_2 \cdot \text{Sc2} - k_1 \cdot \text{Lut1} \\
\frac{d}{dt}\text{Lut2} &= k_1 \cdot \text{Lut1} - k_2 \cdot \text{Lut2} \\
\frac{d}{dt}\text{Lut3} &= k_2 \cdot \text{Lut2} - k_3 \cdot \text{Lut3} \\
\frac{d}{dt}\text{Lut4} &= k_3 \cdot \text{Lut3} - k_4 \cdot \text{Lut4}
\end{aligned} \tag{3.5}$$

$$\text{E2} = e_0 + e_1 \cdot \text{SeF2} + e_2 \cdot \text{PrF} + e_3 \cdot \text{Lut4}$$

$$\text{P4} = p_0 + p_1 \cdot \text{Lut3} + p_2 \cdot \text{Lut4}$$

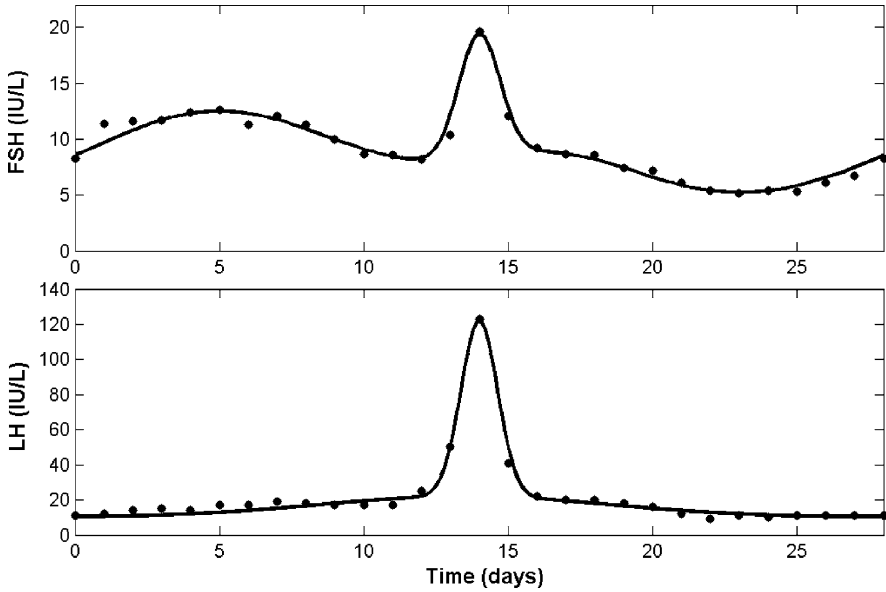
$$\text{IhA} = h_0 + h_1 \cdot \text{PrF} + h_2 \cdot \text{Lut2} + h_3 \cdot \text{Lut3} + h_4 \cdot \text{Lut4}$$

$$\text{IhB} = j_0 + j_1 \cdot \text{PrA2} + j_2 \cdot \text{PrF} + j_3 \cdot \text{OvF} \tag{3.6}$$

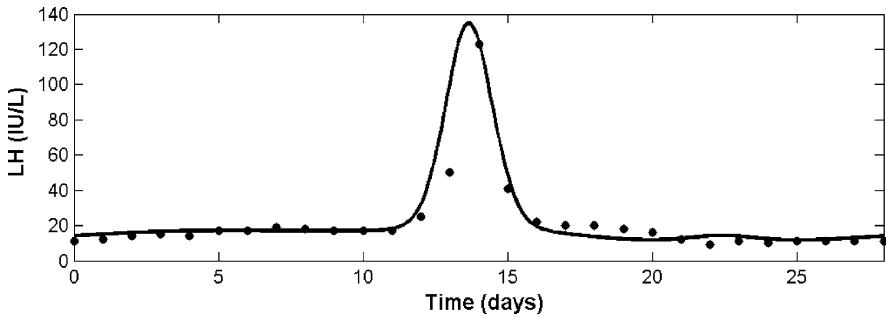
### 3.3 Fitting Parameters

We wished to determine the values of the unknown parameters which made the model best fit the data of [Welt et al. \(1999\)](#). This data set represents the average daily circulating hormone levels for 23 normally cycling women, all aged between 20 and 35 years old. The data were centered to the day of the mid-cycle LH surge. Because the parameter-fitting process was daunting due to the large number of parameters, we began by considering separately the LH system [(3.1) and (3.2)], FSH system [(3.3) and (3.4)], and ovarian system [(3.5) and (3.6)]. We estimated parameters for each of these three systems individually, and then used those results as a starting point for working with a larger, merged model.

The problem this presented is that the three systems are interdependent. The LH system depends on E2 and P4, the FSH system depends on all four ovarian hormones, and the ovarian system depends on LH and FSH. With this in mind, we used explicit functions of time in each system to represent the hormones not being modeled by that system. We used 28-day periodic functions (to match the assumed period of the data from [Welt et al. \(1999\)](#)) consisting of a baseline constant term plus one or more peaks generated by negative-exponential terms. Such terms take the form  $c \cdot \exp(-(t - a)^2)$ , where  $c$  is the amplitude of the peak and  $a$  is the time (in days) at which the peak occurs. As an example, we show in [Fig. 3.5](#)



**Fig. 3.5** Time-dependent input functions for FSH and LH, along with Welt data. Dots represent the clinical data, while lines indicate the approximating functions



**Fig. 3.6** Unmerged LH system model output (line), together with Welt data for LH (dots)

the time-dependent function plots, along with the data, for the two pituitary functions. Additional information regarding the time-dependent functions is included in Sect. 3.3 of [Pasteur \(2008\)](#).

The LH system has 12 parameters, counting the biological constants  $r_{LH}$  and  $\nu$ , while the FSH system has 10 parameters, including  $r_{FSH}$  and the same  $\nu$ . Figures 3.6 and 3.7 show the best fits found for these models. Additional details (including lists of the parameter values) are given in Sects. 3.3 and 5.5, respectively, of [Pasteur \(2008\)](#).

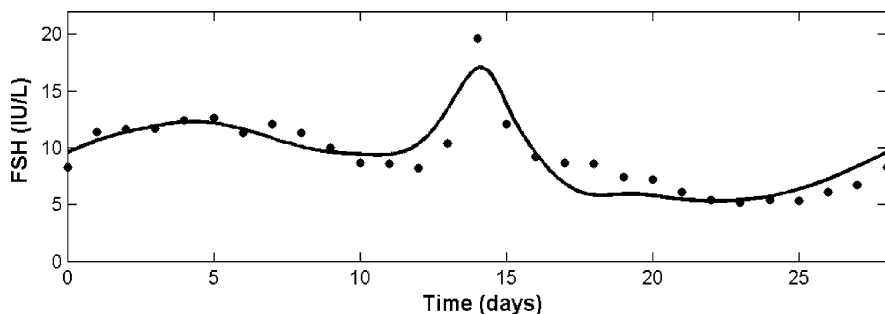


Fig. 3.7 Unmerged FSH system model output, together with Welt data for FSH

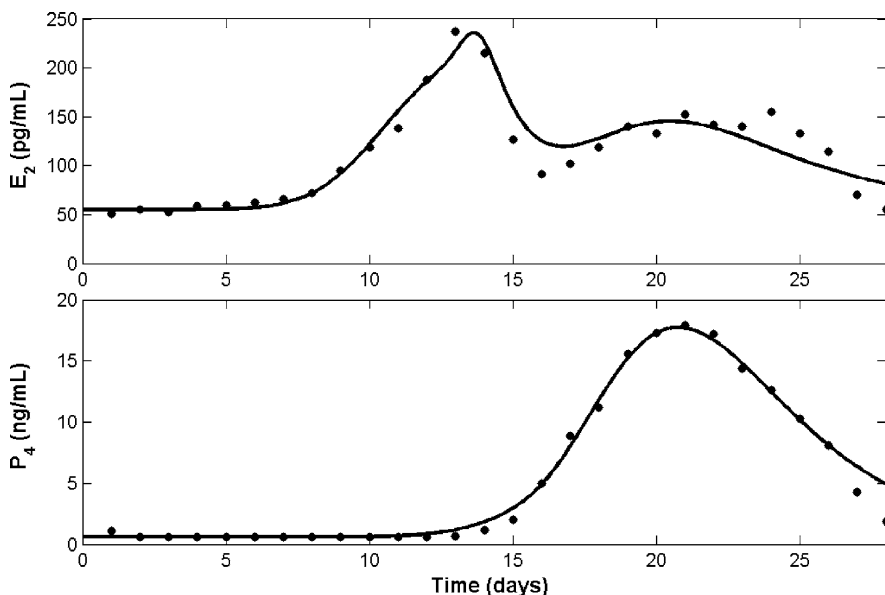


Fig. 3.8 Output for E2 and P4 from the unmerged ovarian model, with Welt data

In the ovarian system, there are 21 parameters for the 12 differential equations (3.5) and 16 additional parameters for the 4 auxiliary equations (3.6). Because the ovarian stages are merely a modeling tool, we include plots only of the auxiliary equations (Figs. 3.8 and 3.9), which show the best fits for the ovarian hormone levels, computed as linear combinations of the 12 ovarian stages, via (3.6).

After obtaining parameter estimates using these three separate models, the next step in the modeling process was to merge them, eliminating the need for time-dependent input functions. Successful completion of this step created a model which is time-autonomous and which can be validated by other data.

The merged model includes 16 differential equations ((3.1)–(3.5)), some with constant delays, plus four auxiliary equations (3.6). There are 58 parameters in this

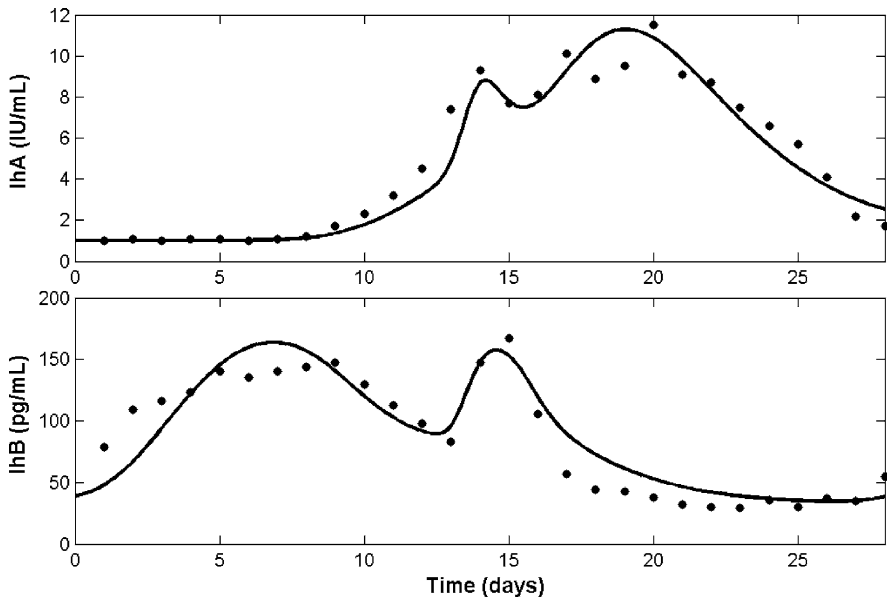


Fig. 3.9 Output for the inhibins from the unmerged ovarian model, with Welt data

merged model, of which only three could be obtained from the physiological literature. Finding the best-fit parameter set was a high-dimensional problem in nonlinear global optimization, because we wished to minimize the total error over the six hormone outputs, as compared to the Welt data, across several cycles. Further complicating matters, the appropriate initial conditions were known for only 2 of the 16 state variables, because neither the releasable pool holdings nor the 12 ovarian stage values are readily approximated at the luteal-follicular transition.

A variety of techniques were used in the parameter identification process, most notably the Nelder–Mead local optimization method, minimizing a weighted total of the sums of squared errors for the six hormone profile outputs. Sensitivity analysis (discussed in Sect. 3.4) was helpful in understanding the effects of changing individual parameters, as well as in determining which parameters have proportionally the largest effects on various output measures.

In Fig. 3.10, we show the model output for the optimized parameter set; the associated values for all 58 parameters are listed in Sect. 5.5 of [Pasteur \(2008\)](#).

These hormone profiles eventually occur for any biologically realistic initial conditions, because there is only one periodic solution and it is an asymptotically stable (i.e., attracting) solution. The period of the solution is 28.0 days, matching both the assumed period of the Welt data and the most common menstrual cycle length ([Speroff et al. 1999](#); [Vollman 1977](#); [Treloar et al. 1967](#)). However, the existence of only one stable periodic solution stands in contrast to the findings of [Clark et al. \(2003\)](#), in which two stable periodic solutions – one normal and one anovulatory – were found after fitting a similar model to a different data set, the data

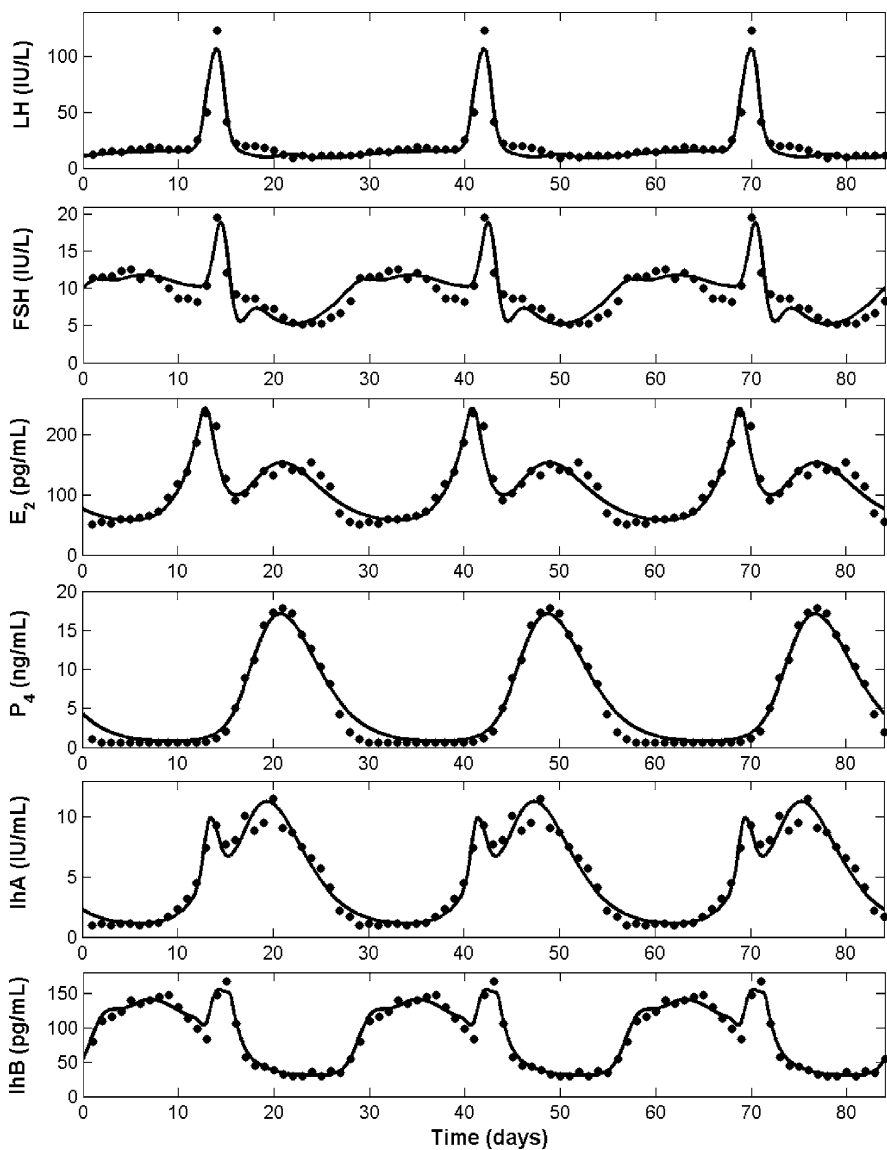


Fig. 3.10 Stable normal cycle for the merged model (*line*), with Welt data (*dots*)

from McLachlan et al. (1990). The McLachlan data set included only total inhibin and had a higher mid-cycle E<sub>2</sub> peak, perhaps explaining the different qualitative result. These two solutions could be thought of as representing two different women, only one of whom has the possibility of sustained abnormal, anovulatory menstrual cycles, in the absence of external effects.

### 3.4 Parameter Sensitivity and Bifurcations

Because of the difficulty of global optimization involving a large number of parameters, together with the complicated form of the differential equations, it is very possible that there are other parameter sets which fit the data as well, or better than the one that we found. The Welt data set consists of average hormone concentrations throughout the menstrual cycle, but clearly there are substantial differences among healthy women. Given that our model is deterministic, individual differences among women can only be expressed using different parameter sets. With this in mind, we wished to consider whether changes to individual parameters could lead to qualitatively different dynamic behavior. In mathematical terms, we were interested in bifurcations, particularly those incurred by changes to a single parameter.

To determine which parameters were most likely to cause substantive changes in the hormone profiles, we computed normalized sensitivity coefficients. These are quantitative measures of the relative effect of a parameter change on some output measure  $X$ . Hence, for a change in some parameter from  $p$  to  $p(1 + \varepsilon)$ , the relative sensitivity coefficient  $S(p)$  is given by (3.7).

$$S(p) = \frac{1}{\varepsilon} \left( \frac{X(p + \varepsilon \cdot p) - X(p)}{X(p)} \right) \quad (3.7)$$

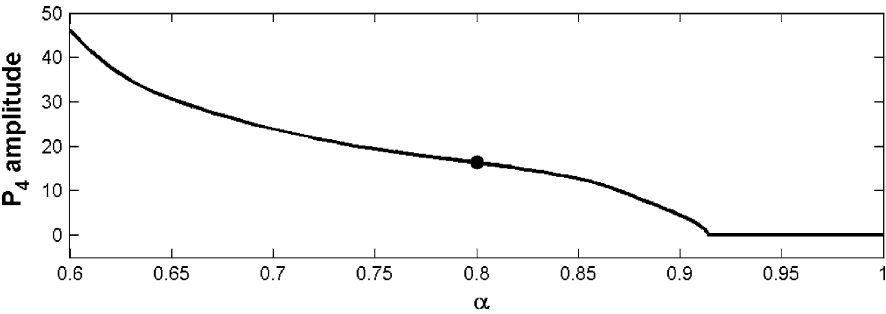
If a large change in a particular parameter is required to bring about a small change in model output, then the associated parameter has a sensitivity coefficient with a very small magnitude. On the other hand, if a small parameter change has substantial impact on the model output, then the coefficient has a large magnitude. The sign of a sensitivity coefficient is determined by whether increasing a parameter causes a measured output to increase or decrease. These coefficients are dimensionless, allowing for straightforward comparisons among all parameters in a model. It is important to note that sensitivity coefficients are tied to the parameter set at which they are measured, and thus will vary for different parameter sets.

Given that our model has six hormone profiles as outputs, there are a variety of measures we could have used in determining sensitivity. Because this process can be easily automated, instead of settling on any one, we computed sensitivity for each parameter with respect to many output measures, typically the peak and minimal levels of each hormone at various points in the cycle, as well as with respect to the cycle period length. Looking through the resulting tables of data, we identified which parameters are the most sensitive overall.

Six parameters stand out as the most sensitive, as shown in Table 3.1, and thus these were the parameters most likely to be associated with qualitative changes in the dynamics of the model. The most sensitive parameter is  $\alpha$ , which is a fractional exponent indicating the strength of the promotion by LH of mass transfer between 2 stages in the ovarian model. Two other sensitive parameters,  $c_1$  and  $c_2$ , are likewise involved in early stages of the ovarian model. Matching reasonable intuition, the threshold values in both Hill functions,  $Km_{LH}$  and  $f_1$ , also have high-magnitude sensitivity coefficients. The baseline LH synthesis coefficient  $v_{0,LH}$  rounds out this

**Table 3.1** *The most sensitive parameters, and their associated sensitivity coefficients, with respect to the mid-cycle E2 peak. Positive coefficients indicate that increasing the parameter leads to a higher E2 peak, while negative coefficients indicate that a parameter increase leads to a lower peak concentration of E2*

| Parameter   | $\alpha$ | $Km_{LH}$ | $c_2$ | $c_1$ | $v_{0,LH}$ | $f_1$ |
|-------------|----------|-----------|-------|-------|------------|-------|
| Sensitivity | -2.37    | 1.16      | -1.03 | 0.87  | -0.60      | 0.59  |



**Fig. 3.11** *Bifurcation diagram for the parameter  $\alpha$ , showing the amplitude of hormone change throughout the cycle, as  $\alpha$  is varied. The dot indicates the best fit value  $\alpha = 0.8001$ . For  $\alpha > 0.913$ , the amplitude is zero, so the system is at equilibrium, i.e., steady-state. This equilibrium is stable and could represent a woman taking a continuous dose of oral contraceptives (discussed in Sect. 3.5) as there would also be no LH surge, and thus no ovulation*

group of six. For the above list, summarized in Table 3.1, we consider sensitivity with respect to the preovulatory peak of E2, which triggers the LH surge, leading to ovulation; however, a variety of other measures could be used, and many of the same parameters would still top the list. Comprehensive tables of sensitivity coefficients are included in Appendix B.2 of [Pasteur \(2008\)](#).

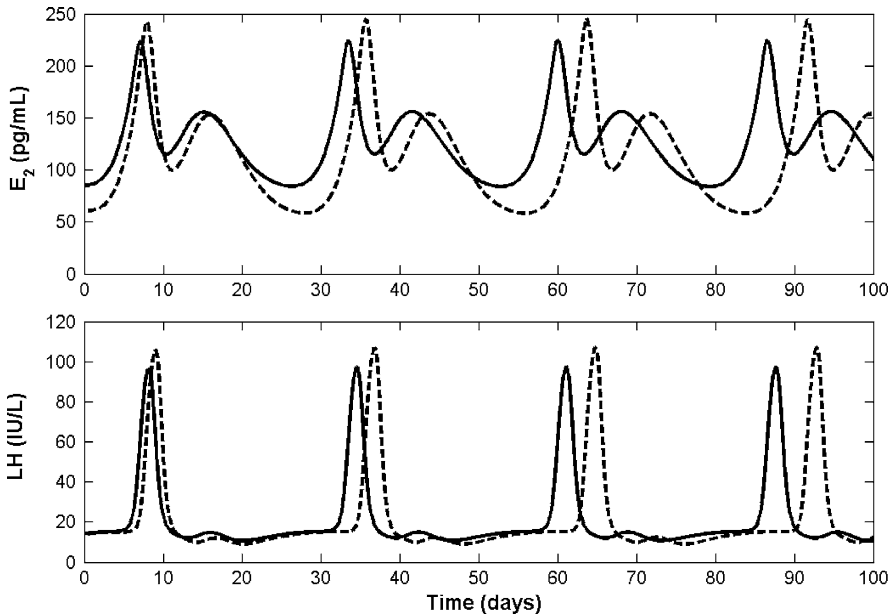
In the model of [Clark et al. \(2003\)](#), two stable periodic solutions were observed for the best-fit parameter set. The implication is that, according to that model, a woman could have either a normal or an abnormal menstrual cycle, sustained over a long duration of time, depending on her initial hormone levels. Furthermore, [Clark et al. \(2003\)](#) showed that a short-term dose of external hormones could cause a change from an abnormal cycle to a normal one, or the reverse. Similar results, for a model different than the one presented here, were shown in Chap. 4 of [Pasteur \(2008\)](#), by changing  $Km_{LH}$ , the E2 threshold for LH synthesis.

In contrast, for our model, we observed at most one stable periodic solution, regardless of varying any single parameter from Table 3.1 within biologically realistic ranges. The existence of a unique stable periodic solution implies that after any disruptions due to external influences, normal menstrual cycles will eventually be resumed. Figure 3.11 shows that by altering the value of the most sensitive parameter  $\alpha$ , we obtained a different type of change in the model behavior, i.e., the periodic solution is replaced by a solution in which hormone levels remain constant over time and ovulation cannot occur.

### 3.5 Exogenous Hormone Effects

Ingestion of exogenous reproductive hormones by women is common today, sometimes intentional (as in pharmaceutical use) and other times not (through pollutants in drinking water). One of the goals of a model such as the one we present is to predict the effects of external hormonal influences. Based on standard contraceptive uses, we know that a sufficiently large daily dose of E2 suppresses ovulation. In our model, we included exogenous E2 by including one or more additional step function terms in the auxiliary (3.6) for E2. As a part of the validation process, we conducted a dose–response analysis of external estrogen intake. The mid-cycle LH surge is the key marker of ovulation; if the LH surge is sufficiently suppressed, ovulation will not occur.

In Figs. 3.12–3.14, we present three cases, involving continuous low, medium, and high doses, respectively, of external E2. In each case, the solid line represents a woman treated with external E2, and the dashed line is the untreated control. In the low-dose case (Fig. 3.12), the bloodstream E2 concentration is raised by 25 pg/ml, an increase of roughly 10% from the peak level observed in our model. There is an associated cycle length reduction from 28.0 to 26.5 days, but only a minimal decrease in the peak amplitude of LH, so ovulation would likely still occur. With a medium dose (75 pg/ml, roughly 30% above the natural peak), LH is strongly suppressed, as shown in Fig. 3.13, and E2 concentrations also take a low-amplitude



**Fig. 3.12** Effects of a 25 pg/ml dose of external E2. The solid curves show the model output with treatment, while the broken curves represent a control group from Fig. 3.10

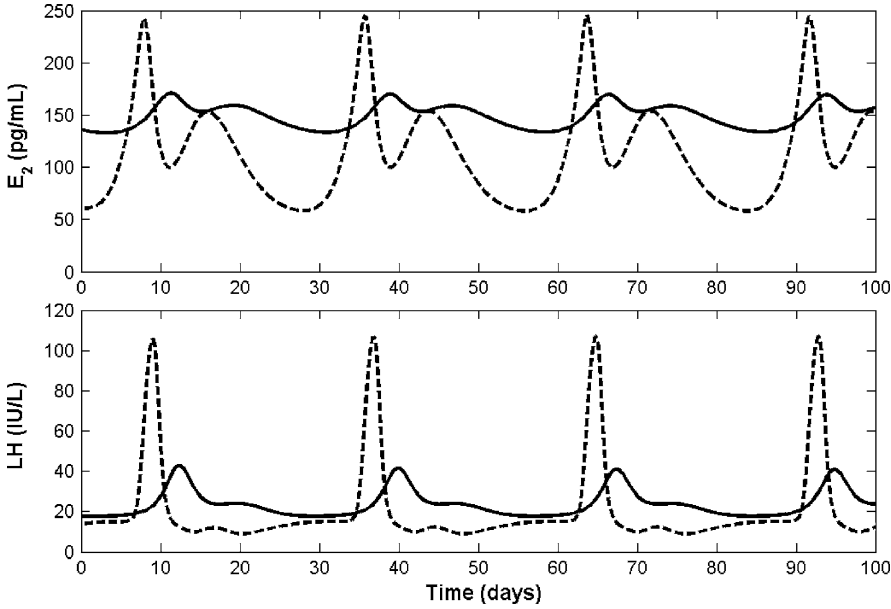


Fig. 3.13 LH surge suppressed by a 75 pg/ml dose of external E2

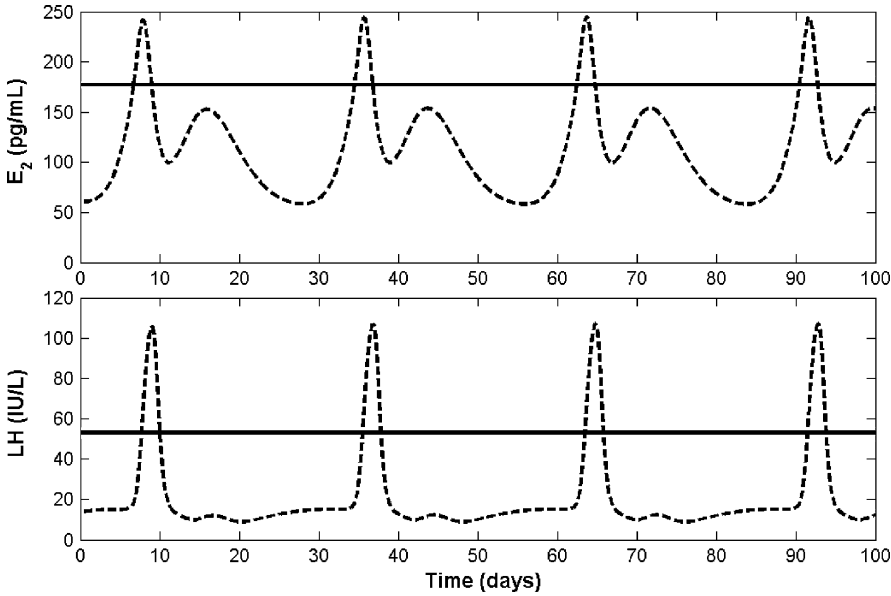


Fig. 3.14 Equilibrium induced by a 125 pg/ml dose of external E2

profile. With a sufficiently large dose (125 pg/ml, about half as much as the observed peak in the Welt data), the model predicts a pharmaceutically induced equilibrium, with constant levels of all hormones, as shown in Fig. 3.14.

All of this behavior matched our expectations based on standard oral contraceptive treatment protocols. Increasingly large amounts of exogenous E2 suppress the LH surge (and hence, suppress ovulation) to an increasingly large degree. It is also noteworthy that upon cessation of the treatment, the modeled hormone concentrations return to the stable normal cycle of Fig. 3.10 after a few months. These unsurprising results help validate the model.

### 3.6 Conclusion

We have presented a model for hormonal control of the female reproductive endocrine system which produces, with reasonable accuracy, hormone profiles typical of a young, healthy woman, throughout several consecutive menstrual cycles. Unlike other prior models such as Clark et al. (2003), this model takes advantage of data from relatively new bioassays for two forms of inhibin. Intuitively reasonable effects on LH have been shown for the treatment with external E2. The use of modern high-speed computers and specialized software tools has allowed for simulation and analysis with a complex model.

However, further improvements can be made to the model presented. The FSH profile in Fig. 3.10 is a far-from-perfect fit, particularly in the week surrounding ovulation in each cycle. Additionally, simulations of pharmaceutical use of external P4 and combinations of E2 and P4, both of which are used clinically for contraception in some cases (Speroff et al. 1999), did not result in the expected suppression of the LH surge. These issues demonstrate a need for refitting the parameter set, perhaps with one or more global optimization techniques, to obtain a model which better reflects the biological reality. It is possible that the underlying equations, not just the parameters, will have to be altered to achieve this goal. It is unknown whether the effect of E2 on LH is on the synthesis of LH or the release. Our present model assumes that E2 inhibits LH release but, at high levels, promotes LH synthesis. Future work will investigate the possibility that the effect on LH release changes from inhibitory to stimulatory after E2 reaches a threshold level for an extended period of time.

Additionally, a deterministic model is limited in scope, because each woman has a different hormone profile. There are also fluctuations over time (from one cycle to the next) in the hormone concentrations of the same woman. Both of these issues point to the need for a stochastic model. Such a model, based on randomized parameters, could show individual variations across the population being studied, and may even allow for simulated clinical trials of hormonal treatment regimens.

## References

- D. T. Baird. The ovary. In C. R. Austin and R. V. Short, editors, *Hormonal control of reproduction*, pages 91–114, Cambridge University Press, Cambridge, 1984
- R. J. Bogumil, M. Ferin, J. Rootenberg, L. Speroff, and R. L. Vande Wiele. Mathematical studies of the human menstrual cycle. I: Formulation of a mathematical model. *Journal of Clinical Endocrinology and Metabolism*, 35:126–143, 1972
- L. H. Clark, P. M. Schlosser, and J. F. Selgrade. Multiple stable periodic solutions in a model for hormonal control of the menstrual cycle. *Bulletin of Mathematical Biology*, 65:157–173, 2003
- Y. D. Coble, P. O. Kohler, C. M. Cargille, and G. T. Ross. Production rates and metabolic clearance rates of human follicle-stimulating hormone in premenopausal and postmenopausal women. *The Journal of Clinical Investigation*, 48:359–363, 1969
- D. L. Davis, H. L. Bradlow, M. Wolff, T. Woodruff, D. G. Hoel, and H. Anton-Culver. Medical hypothesis: Xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives*, 101:372–377, 1993
- N. P. Groome, P. J. Illingworth, M. O’Brien, R. Pai, F. E. Rodger, J. P. Mather, and A. S. McNeilly. Measurement of dimeric inhibin B throughout the human menstrual cycle. *Journal of Clinical Endocrinology and Metabolism*, 81:1401–1405, 1996
- M. E. Hadley. *Endocrinology*. Simon and Schuster Co., Englewood Cliffs, third edition, 1992
- J. Keener and J. Sneyd. *Mathematical physiology I: Cellular physiology*. Springer, New York, second edition, 2009
- P. O. Kohler, G. T. Ross, and W. D. Odell. Metabolic clearance and production rates of human luteinizing hormone in pre- and postmenopausal women. *The Journal of Clinical Investigation*, 47:38–47, 1968
- J. H. Liu and S. S. C. Yen. Induction of midcycle gonadotropin surge by ovarian steroids in women: A critical evaluation. *Journal of Clinical Endocrinology and Metabolism*, 57:797–802, 1983
- R. I. McLachlan, N. L. Cohen, K. D. Dahl, W. J. Bremner, and M. R. Soules. Serum inhibin levels during the periovulatory interval in normal women: Relationships with sex steroid and gonadotropin levels. *Clinical Endocrinology*, 32:39–48, 1990
- S. Muttukrishna, T. Child, G. M. Lockwood, N. P. Groome, D. H. Barlow, and W. L. Ledger. Serum concentrations of dimeric inhibins, activin A, gonadotropins, and ovarian steroids during the menstrual cycle in older women. *Human Reproduction*, 15:549–556, 2000
- K. Oktay, H. Newton, J. Mullan, and R. G. Gosden. Development of human primordial follicles to antral stages in SCID/hpg mice stimulated with follicle stimulating hormone. *Human Reproduction*, 13:1133–1138, 1998
- T. D. Pache, J. W. Wladimiroff, F. H. de Jong, W. C. Hop, and B. C. J. M. Fauser. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. *Fertility and Sterility*, 54: 638–642, 1990
- R. D. Pasteur. *A multiple-inhibin model of the human menstrual cycle*. PhD thesis, North Carolina State University, 2008. URL <http://www.lib.ncsu.edu/theses/available/etd-06102008-194807/>
- I. Reinecke and P. Deuflhard. A complex mathematical model of the human menstrual cycle. *Journal of Theoretical Biology*, 247:303–330, 2007
- R. A. Rudel, S. J. Melly, P. W. Geno, G. Sun, and J. G. Brody. Identification of alkylphenols and other estrogenic phenolic compounds in wastewater, septage, and groundwater on Cape Cod, Massachusetts. *Environmental Science and Technology*, 32:861–869, 1998
- P. M. Schlosser and J. F. Selgrade. A model of gonadotropin regulation during the menstrual cycle in women: Qualitative features. *Environmental Health Perspectives*, 108:873–881, 2000
- J. F. Selgrade and P. M. Schlosser. A model for the production of ovarian hormones during the menstrual cycle. *Fields Institute Communications*, 21:429–446, 1999
- L. Speroff, R. H. Glass, and N. G. Kase. *Clinical gynecologic endocrinology and infertility*. Lippincott Williams and Wilkins, Philadelphia, sixth edition, 1999

- A. E. Treloar, R. E. Boynton, G. B. Borghild, and B. W. Brown. Variation of the human menstrual cycle through reproductive life. *International Journal of Fertility*, 12:77–126, 1967
- R. F. Vollman. The menstrual cycle. In E. Friedman, editor, *Major problems in obstetrics and gynecology*, pages 1–193, 1977
- C. K. Welt, K. A. Martin, A. E. Taylor, G. M. Lambert-Messerlian, W. F. Jr. Crowley, J. A. Smith, D. A. Schoenfield, and J. E. Hall. Frequency modulation of follicle-stimulating hormone (FSH) during the luteal-follicular transition: Evidence of FSH control of inhibin B in normal women. *Journal of Clinical Endocrinology and Metabolism*, 82:2645–2652, 1997
- C. K. Welt, D. J. McNicholl, A. E. Taylor, and J. E. Hall. Female reproductive aging is marked by decreased secretion of dimeric inhibin. *Journal of Clinical Endocrinology and Metabolism*, 84: 105–111, 1999
- S. S. C. Yen. The human menstrual cycle: Neuroendocrine regulation. In S. S. C. Yen, R. B. Jaffe, and R. L. Barbieri, editors, *Reproductive endocrinology: Physiology, pathophysiology, and clinical management*, pages 191–217, W. B. Saunders Co., Philadelphia, 1999
- J. R. Young and R. B. Jaffe. Strength-duration characteristics of estrogen effects on gonadotropin response to gonadotropin-releasing hormone in women. *Journal of Clinical Endocrinology and Metabolism*, 42:432–442, 1976
- A. J. Zeleznik and C. R. Pohl. Control of follicular development, corpus luteum function, the maternal recognition of pregnancy, and the neuroendocrine regulation of the menstrual cycle in higher primates. In J. D. Neill, editor, *Knobil and Neill's physiology of reproduction*. Elsevier, 2006

# Chapter 4

## Modeling Transport Processes and Their Implications for Chemical Disposition and Action

Nick Plant

### 4.1 Introduction

The body is constantly being exposed to chemicals: This exposure may range from deliberate (e.g., therapeutic agents, food chemicals) to undesirable (e.g., environmental and food contaminants). The first role of the body must therefore be to sort these chemicals, allowing ingress of those that are beneficial to the body while rapidly removing those that could elicit harm. This process must be achieved in a rapid and efficient manner and be coordinated such that the most efficient response is elicited for any given chemical exposure. To be able to model such behavior, and hence predict the outcome of any subsequent exposure, it is first necessary to understand the basic mechanisms by which the body responds to chemical insult.

#### 4.1.1 *The Fate of Chemicals in the Body*

The fate of a chemical within the body is determined by the processes of absorption, distribution, metabolism and excretion (ADME). These four stages control not only the amount of any given chemical that enters the body but also the rate at which it is subsequently chemically altered and excreted from the body (Plant 2003). These four processes are outlined in Fig. 4.1 and detailed in the following sections.

##### 4.1.1.1 Absorption

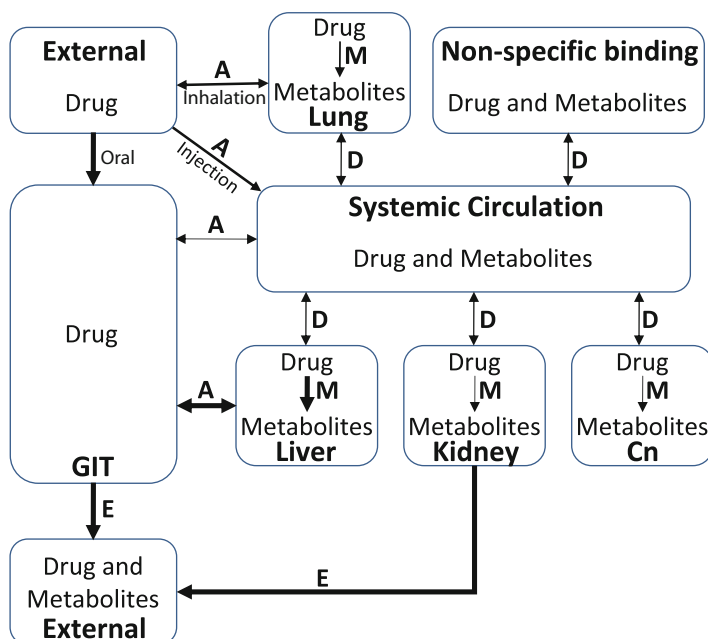
The body is essentially a series of aqueous environments (cytoplasm) bounded by lipid (membranes). Hence, to enter, and subsequently distribute, around the body,

---

N. Plant (✉)

Centre for Toxicology, Faculty of Health and Medical Sciences, University of Surrey,  
Guildford, Surrey GU2 7XH, UK

e-mail: [n.plant@surrey.ac.uk](mailto:n.plant@surrey.ac.uk)



**Fig. 4.1** Outline of ADME processes within the body. From the external of the body a chemical may undergo a number of processes before its final excretion to the external environment. Arrows indicate potential chemical movement, with *thickness* indicating propensity for a particular route to occur. A Absorption, D Distribution, M Metabolism, E Excretion, Cn Other compartments within the body, GIT Gastrointestinal tract

chemicals must be able to cross lipid membranes. Due to this, those chemicals that are best absorbed tend to be lipophilic in nature, meaning that they can cross these membrane barriers by simple passive diffusion. Such a process follows simple first-order kinetics<sup>1</sup> and can be therefore modeled, at the simplest level, by simple mass action.<sup>2</sup> However, for those chemicals that are too hydrophilic to be efficiently absorbed across lipid membranes two potential entry routes exist. First, the use of protein “tunnels”, through which these chemicals can diffuse, while having their hydrophilic nature shielded from the lipid-based membrane. Second, energy can be expended to move chemicals against a concentration gradient; this energy can be either in the direct form of hydrolysis of ATP to ADP or indirectly using cotransport or antitransport of chemicals down gradients, thus providing a motive force. Transport processes that utilize a transmembrane protein during their functioning, whether it is for facilitated diffusion or active transport, are many and varied in

<sup>1</sup> First-order kinetics is defined as any kinetic reaction where the rate of the reaction is not limited by the level of catalyzing protein, i.e., the protein is not saturated.

<sup>2</sup> Mass action refers to the determination of reaction rate as the product of substrate concentration and the rate constant for the reaction ( $v = [S] \times k$ ).

nature. These processes act as both influx and efflux transporters, but all essentially can be defined as saturable processes exhibiting both first- and zero-order kinetics<sup>3</sup> (Plant 2003).

#### 4.1.1.2 Distribution

The process of chemical distribution around the body is essentially driven by the blood flow to any given tissue, although it can be modulated by a number of other factors. First, nonspecific binding of chemicals to plasma proteins retains them in the systemic blood supply, delaying both their distribution to organs and their excretion from the body (Gibson and Skett 2001): Depending upon the nature of this nonspecific binding it can be described as either an equilibrium constant ( $K_{eq}$ ), being the ratio of the forward and reverse rate constants, or have terms describing the saturable nature of the binding. Second, as has been described for the absorption phase of ADME, the presence of transport proteins in a given tissue can regulate chemical distribution to that tissue, either causing selective uptake or exclusion (Plant 2003). This selective distribution to tissues is a central mechanism used to protect organs susceptible to toxic damage, such as the brain, and to concentrate chemicals in metabolically active organs such as the liver to enable their further processing (Ayrton and Morgan 2001).

#### 4.1.1.3 Metabolism

Alteration of a chemical structure has two major effects: First, as the routes of excretion are water-based it is often necessary to alter lipophilic chemicals, which are well absorbed, into more hydrophilic chemicals that can be efficiently excreted. Second, the act of metabolism alters a chemical structure, often changing its chemical reactivity. Hence, a potential protective mechanism is the rapid deactivation of chemicals before they can cause toxic damage to a tissue.

Metabolism is classically divided into two phases, catalyzed by distinct groups of enzymes. Phase I metabolism is predominantly undertaken by the Cytochrome P450 super family of enzymes, while Phase II metabolism is shared by several, smaller, enzyme families, each of which acts to add a specific chemical moiety to a substrate. In both phases, the enzymes show marked substrate promiscuity, producing an adaptive system that can respond to a wide variety of chemical exposures. In general, Phase I reactions act to make substrates more chemically reactive (bioactivation), producing reactive centers for the deactivating reactions catalyzed by Phase II enzymes (Gibson and Skett 2001). It should be noted that Phase I and II do not have to occur sequentially, for if a reactive center already exists within a substrate then this can negate the requirement for Phase I metabolism, as is seen with acetaminophen

---

<sup>3</sup> Zero-order kinetics is defined as any kinetic reaction where the rate of the reaction is limited by the level of catalyzing protein, i.e., where the reaction is saturated.

(Plant 2003). Both Phase I and Phase II reactions are simple enzyme-catalyzed reactions and can thus be described using suitable, saturable, kinetic parameters, such as the Michaelis–Menten equation and its derivatives.

#### 4.1.1.4 Excretion

The final process in the life-cycle of a chemical within the body is its excretion from the organism. This process predominantly occurs via the kidneys, producing urine, and liver, producing feces, although other, minor, routes do exist (Gibson and Skett 2001). The process of excretion occurs in a fundamentally similar manner to absorption, essentially being the reverse, and hence can be modeled as a combination of passive and active processes.

### 4.1.2 *Chemical and Pathophysiological-Mediated Alterations in Drug Disposition*

One central feature of biological systems is their ability to adapt to the environment that they exist within. This feature is especially important within ADME processes as it allows the body to efficiently respond to fluctuations within the environment, thus bringing the body back to homeostasis. To understand the role of this adaptation within biological systems it is important to understand both the concept of robustness in general and with specific relation to ADME, and these two areas will be covered below. For the purpose of this text, the term network will be used to describe the interactions of a number of chemical and biological species, including drugs, proteins, and genes, within the cellular environment.

Within biological systems there is a requirement for robustness, defined as the ability of that biological system to continue to carry out its fundamental tasks (Kitano 2004). Perfect robustness within a network is unlikely to occur unless the network fulfills some very specific criteria (Shinar et al. 2007), and indeed such perfect robustness is probably relatively rare in biological systems. However, the use of feedback mechanisms allows systems to adapt to alterations in environmental conditions, thus achieving quasirobustness. Within ADME the requirement for robustness is obvious, as chemical exposure is, by definition, a move away from homeostasis; without biological robustness body functioning would rapidly breakdown. In addition, it should be noted that ADME processes also control the fate of endogenous chemicals within the body; any alteration in ADME processes caused in response to external chemicals will potentially impact upon these endogenous processes and this can result in toxicity (Plant 2004). In essence, without robustness the ADME system that is central to body function would not be able to respond effectively to chemical exposure from the environment, while still maintaining its endogenous biological roles. This robustness is achieved through three, interconnected systems. First, promiscuity within the ligand specificity for transporters and drug

metabolizing enzymes ensures that for any given chemical, whether it is endogenous or exogenous in origin, several complimentary systems can mediate the efficient removal of the stimulating chemical (Plant 2004, 2007; Watkins et al. 2001; Smirlis et al. 2001). Second, these ADME pathways are subject to complex feedback and feed-forward loops, coordinating protein expression levels with their requirement to handle stimulating chemicals (Plant 2007; Pascussi et al. 2000a, b; Aouabdi et al. 2006). Third, activation of ADME pathways is closely coupled with the activation of protective mechanisms that remove toxic damage elicited by stimulating chemicals should it occur before the chemical can be safely removed (Plant 2003; Roberts et al. 1997).

In addition to alterations in ADME elicited by chemical exposure, there is also a well-described alteration in microenvironment caused by pathophysiology, with the most obvious example being during tumorigenesis. The alterations in gene and protein expression in tumors have been well described, initially at the level of the tumor in general (Kim et al. 2009; Sotiriou and Pusztai 2009), providing a framework to both aid in the grading of tumors (Ramaswamy et al. 2001) and predict the response to therapeutic intervention (Villeneuve et al. 2006). Recent work has focused on the heterogeneity of expression within single cells within a tumor (Slack et al. 2008), leading to the potential to model tumor responses at the single cell level. One concept that is now clear is that a central feature of tumor cells is their altered ADME processes, leading to altered handling of chemicals compared to normal tissue, which may result in altered therapeutic efficacy. It is vital to understand this variability in order to be able to optimize therapeutic intervention, either through the development of novel therapeutics, potentially targeted at fragile nodes within a network, the modulation of response networks through coadministration of multiple therapeutics or the optimization of current therapeutic strategies through the increased understanding of these response networks.

### ***4.1.3 Extrapolation of Data Between Biological Scenarios***

For the safe development of novel chemical entities, the optimal usage of existing therapeutics, and the robust risk assessment of human exposure to potential toxicants, it is important to be able to extrapolate data from one scenario to another. Such extrapolation may be between normal and pathophysiological conditions, from individual to population exposures, including the impact of genetic variability, or from model species/test systems to the *in vivo* human situation. Making such decisions earlier in new chemical entity discovery/development is central to ensuring that the most likely leads are progressed down the development pipeline, thus enhancing the possibility of them making it to market. Such decisions are obviously dependent on the ability to make robust extrapolations, and as such the development of novel *in silico* models may represent an important driver in the continued optimization/understanding of human response/safety to drug treatment. In addition, creation of such models will provide important insights into the normal physiological processes that occur in humans.

Given that the chemical industry, both pharmaceutical and industrial, have been undertaking extrapolation of data between scenarios for many decades, how then can systems modeling help? To answer this question, it is necessary to understand the current paradigm on which this extrapolation is based and thus see how this can be improved and/or modified. The next section will cover the traditional approach of pharmacokinetics, while Sect. 4.3 will detail a bottom-up systems approach to ADME modeling.

## 4.2 Traditional Pharmacokinetic Approaches to Modeling Drug Disposition

As detailed in the previous section, there is a need to be able to robustly model the fate of chemicals within the human body. Such mathematical treatment of chemical fate has been undertaken for many decades and is the basis of the science of pharmacokinetics. In traditional pharmacokinetics a reductionist approach is taken, whereby several biological reactions are described by only a single term (Tozer and Rowland 2006). In fact, the underlying principal of traditional pharmacokinetics is the generation of compartments that represent tissues with similar kinetic profiles, as opposed to identifying all the individual reactions within each tissue.

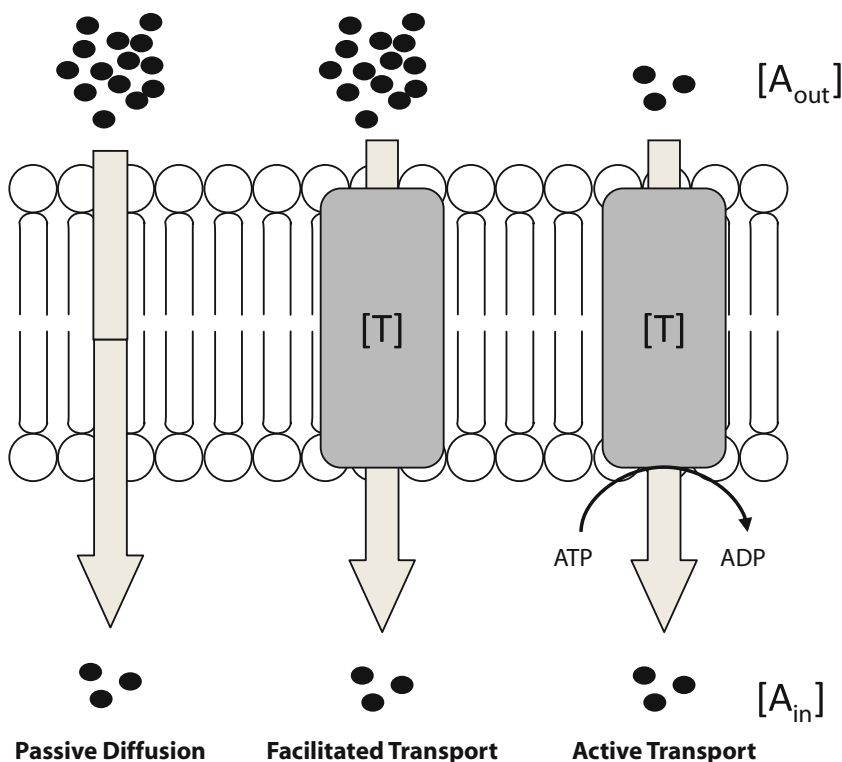
Such an approach has the distinct advantage that models can be constructed from relatively poor data sets, perhaps only containing values for input and/or output levels of the chemical, plus some major transition states in between. Systems such as SimCYP (Rostami-Hodjegan and Tucker 2007) are able to not only model chemical fate within a system but also to allow manipulation of individual parameters such that population responses can be studied. There is, however, an obvious problem with a reductionist-heavy approach, which is that it may potentially limit the ability to examine fine detail within the network. For example, the role of individual parameters, such as the concentration of a single protein, cannot always be studied as the models are based at a detail level above this. Whereas this may not necessarily be an issue for the study of population-based kinetics, it can have potential issues under two circumstances. First, extrapolation of data from one species to another may be seriously skewed if there is not a clear relationship between the model species. For example, many ADME proteins show species-specific characteristics, with orthologues often showing distinct expression levels and kinetic parameters. In addition, altered ADME between species also exists, with, for example, mice and hamsters being highly sensitive to the toxic effects of acetaminophen whereas rabbits and guinea pigs are relatively insensitive (Bessemers and Vermeulen 2001). Second, as discussed in Sect. 4.1.3 there are known differences between normo- and patho-physiological tissue states, and hence it is important to understand how the alteration of these response networks impacts upon biological functioning and response to therapeutic intervention.

### 4.3 More Complex Models of Drug Movement Across Biological Membranes

#### 4.3.1 The Measurement of Chemical Movement Across Biological Membranes

As described in Sect. 4.1, any chemical that moves across a lipid membrane will do so through one of two major routes, direct movement through the lipid bilayer or transfer through an intermediate protein structure (a drug transporter) (Fig. 4.2).

It should be noted that, in reality, the majority of chemicals are transported by a mixture of passive and active transport processes, and hence both modes will need to be modeled for any given chemical. In addition, both of these modes of



**Fig. 4.2** Major potential modes of transport of chemicals across a biological membrane. Nonpolar (lipophilic) drugs cross membranes by simple passive diffusion down their concentration gradients. However, as hydrophilicity increases, protein ( $T$ ) is required to act as a pore for the drug, masking polar groups that would impede transport (facilitated transport). Finally, for transport lacking a concentration gradient, energy must be supplied, often via the hydrolysis of ATP to ADP (active transport)

transport can occur in either direction across the membrane, although in the case of active transport this is dependent upon the involvement of separate influx and efflux transporters.

On the surface it would thus seem relatively simple to model these two modes of transport, with passive and facilitated diffusion being modeled by simple mass action kinetics (4.1), whereas active transport can be modeled as an irreversible Michaelis–Menten style equation (4.2), with the inclusion of a Hill slope if necessary to allow for allosteric interactions (4.3).

$$\frac{\delta A_{\text{out}}}{\delta t} = k \cdot [A_{\text{out}}] - k \cdot [A_{\text{in}}] , \quad (4.1)$$

$$\frac{\delta A_{\text{out}}}{\delta t} = \frac{V_{\text{max}} \cdot [T] \cdot [A_{\text{out}}]}{K_{\text{m}} + [A_{\text{out}}]} , \quad (4.2)$$

$$\frac{\delta A_{\text{out}}}{\delta t} = \frac{V_{\text{max}} \cdot [T] \cdot [A_{\text{out}}]^h}{K_{\text{m}} + [A_{\text{out}}]} , \quad (4.3)$$

where

$A_{\text{in}}$  and  $A_{\text{out}}$  are the concentrations of drug in donor and acceptor compartments (M);  $k$  is the rate constant for drug movement,  $V_{\text{max}}$  is the maximal rate of drug transport across membrane (moles/min/mg protein),  $K_{\text{m}}$  is the drug concentration required to achieve a transport rate equal to  $1/2 V_{\text{max}}$  (M),  $h$  is the hill slope,  $t$  is the time in minutes, and  $[T]$  is the concentration of drug transporter (M).

However, as will be discussed in the following section, such simplistic equations do not fully encompass the complexity of membrane transport, with refinements needed to increase the accuracy of modeling. In addition, it is worth noting that the experimental system used to gain kinetic data can have a major impact on the in vivo relevance of this data; it is hence worthwhile first addressing the general assumptions often made during kinetic modeling of transport processes.

### 4.3.2 *General Considerations for Measuring Movement of Drugs Across Biological Membranes*

It is an obvious statement that when kinetic data is derived from in vitro systems for use in in silico models, it is important that the data accurately reflects the biological model that is being simulated. However, whereas in vitro systems provide good surrogates for in vivo, it is important to note the potential differences between the two scenarios as this may help explain why derived models do not achieve full predictivity. These differences are well reviewed by [Youdim et al. \(2003\)](#), with a brief description given below.

### 4.3.2.1 Simple Versus Complex Measurement Systems

As will be detailed within the next section, transport across biological membranes can be essentially divided into the passive and active transport components of the total transport. As such a number of different systems have been developed to measure these two compartments, including *in silico* assessment of  $\log P^4$  or  $\log D^5$  (Sawada et al. 1999); use of preformed lipid vesicles (Zhou et al. 2009); cell lines expressing individual transport proteins (Acharya et al. 2006); individual cell lines (Walle and Walle 1998); coculture systems (Perriere et al. 2007); and *in vivo* measurements (Vlaming et al. 2009). As these systems range from the highly simplistic, but easy to handle, to complete *in vivo* systems where many confounding factors to accurate measurement exist, then it is important to consider the desired outcomes before a test system is chosen. The choice of system will impact on not only the general assumptions outlined below but also the level of detail generated within the study and the manner in which this needs to be treated.

### 4.3.2.2 Ionization Status of the Drug

A central role of drug transport processes is the regulation of chemical access from the environment; as such a major site of drug transport is the gastro-intestinal tract (GIT). The GIT is a complex tissue, with many functional subdivisions along its length, essentially producing microenvironments that need to be considered separately. For example, there is considerable variance in pH along the length of the tract, with values varying from alkaline to acidic. In accordance with the Henderson–Hasselbalch equation (4.4), the ionization state of a chemical is dependent on the pH of the local environment. Such variation may have a considerable impact on passive diffusion of chemicals given that the pH partition hypothesis states that only the nonpolar form of a chemical will cross a lipid membrane by passive diffusion. The impact of local pH on the permeability of a membrane to a chemical is often referred to as the passive permeability  $P$ .

$$\text{pH} = \text{p}K_a + \frac{\log[A^- + H^+]}{[AH]}, \quad (4.4)$$

where

$A^-$  is the polar drug form concentration (M),  $H^+$  is the proton concentration (M),  $AH$  is the nonpolar drug form concentration (M), and  $\text{p}K_a$  is  $\log_{10}$  (Dissociation constant for  $AH$ ).

---

<sup>4</sup>  $\log P$  is defined as the log 10 of the [chemical in lipid (usually octanol)]/[chemical in water] at a specified pH.

<sup>5</sup>  $\log P$  is defined as the log 10 of the [chemical in lipid (usually octanol)]/[chemical in water] at pH 7.4.

It is important to note that it is the pH of the microenvironment bordering the membrane that is important, which may not necessarily relate to the pH of the entire lumen. For example, whereas the pH of the stomach is highly acidic ( $\text{pH} \approx 2$ ), the microenvironment of the membrane tends toward neutral, on average pH 8 (Rechkemmer 1991). This variation in pH can have a large impact on  $P$ , as demonstrated by Palm et al. (1999) who examined the effect of altering the degree of ionization of alfentanil and cimetidine from 5 to 95% on the passive diffusion of these chemicals. They demonstrated that transport of the nonpolar form was 150- to 30-fold more rapid than the ionized drug, respectively. In addition, when the fraction of nonpolar drug ( $f_u$ ) was less than 0.1 the contribution of ionization status to determining  $P$  became significant. If one considers the pH values observed in the gut membrane microenvironment, then  $f_u$  was consistently less than 0.1, meaning that ionization status is likely to have a significant impact in the biological systems. Indeed, Palm et al. (1999) demonstrated that pH variability within the physiological range altered  $P$  by as much as 2.5-fold.

#### 4.3.2.3 Heterogeneity in Drug Dispersion

It is generally presumed that chemical is homogeneously distributed throughout the liquid phase; however, this is certainly not the case close to biological membranes, as the area either side of the membrane forms an “unstirred water layer” (UWL), through which the chemical must diffuse prior to entering the membrane. The size of the UWL can be altered through vigorous stirring of the medium in cell culture systems, but never totally removed, and in the case of transport assays the UWL may exceed 1 mm; in comparison, in vivo the UWL is estimated to be 30–100  $\mu\text{m}$  thick within the GIT (Lennernas 2007). Given that most transport assays do not routinely utilize medium agitation then it can be presumed that the UWL in these situations is larger than expected in vivo. Such an experimental set-up will, of course, reduce the robustness of any model based upon in vitro data for predicting in vivo scenarios.

#### 4.3.2.4 Chemical Sequestration

The majority of in vitro techniques for measuring chemical movement across biological membranes function by measuring the amount of chemical in either compartment or in both. However, it should be noted that the ability to cross a membrane is determined by the free-fraction of chemical, with chemical being potentially sequestered in to either compartment or even the membrane itself. Nonspecific binding to extra- or intracellular proteins can be easily measured and incorporated into the diffusion equation as an extra term. In addition, as these bindings are reversible then it should be remembered that the total amount of chemical bound will alter with time as free chemical diffuses from one compartment to the next. With regard to the extrapolation from in silico models to in vivo then intracellular binding is assumed to be roughly equal for any given cell type and chemical. However, the

extracellular binding will be determined by the level of binding proteins present in the serum and medium, and data should be corrected to compensate for inequalities in this. One easy solution is to ensure the use of cell culture medium that contains plasma-binding proteins such as albumin, mimicking sequestration in the systemic circulation.

A more critical component is that most chemicals subject to significant passive diffusion will, by definition, be lipophilic in nature. As such, there is the potential for sequestration within the membrane itself. The degree of sequestration is dependent on both the lipophilicity of the chemicals and also the availability of acceptor chemicals within the receiving volume, which are necessary for the extraction of the chemical from the membrane. Experimental estimates for the impact of this retention vary; in Caco-2 cells [Wils et al. \(1994\)](#) estimated retention at only approximately 40% for chemicals with LogD values in the range 3.5–5.2, whereas [Sawada et al. \(1999\)](#) observed up to 89% retention for highly lipophilic chemicals (logP 1.1–15) being tested in MDCK cells.

#### 4.3.2.5 Physico-Chemical Characteristics of the Chemical

It is generally assumed that within the experimental system chemical characteristics remain constant throughout the time of the assay. Indeed, this is often a determined prerequisite for “probe chemicals” used within in vitro test systems (i.e., those chemicals determined to be satisfactory markers for an individual biological event). There are actually three assumptions made within this larger assumption: First, that the chemical is chemically stable over the time of the assay; second, that the chemical is metabolically stable over the time of the assay; third, that the chemical maintains the same physico-chemical characteristics over the concentration range tested within the assay. In general, the first assumption should be relatively robust unless the assay is undertaken over an extremely extended period of time or the initial choice of probe chemical was poor. The second assumption is not always valid, and with the majority of chemicals likely to show some metabolism during the period of the assay. Such loss can be dealt with by either measuring parameters that are resistant to such loss, such as the depletion of chemical from medium where no metabolic activity is present, or through the incorporation of a term to measure the rate of loss via metabolism. As an aside, it should be noted that many cell lines actually have markedly reduced metabolic capacities, compared to in vivo, which is often considered a limiting factor in their utility, as they are unable to mimic the metabolism seen in vivo; however, in the case of maintaining probe chemical integrity this is actually an advantage, as chemical loss through metabolism will be significantly reduced.

The final assumption often made is that the chemical reacts with its environment in an identical manner at all concentration used. However, at higher chemical concentrations most chemicals will pass a “critical micellar concentration” (CMC), above which they form micelle structures that have an increased passive permeability through the membrane. Hence, measurements taken above the CMC will be

unreliable as the ratio of passive to active transport will be substantially altered. The solution to this is simple; identify the CMC for the probe chemical before testing transport rates and do not exceed it within the test system.

#### 4.3.2.6 ATP Usage Within the Test System

Many active transport processes require the hydrolysis of ATP to provide the energy for transport. Hence, it could be envisaged that ATP levels and kinetics may influence the overall transport efficiency. In general, however, the input of ATP into active transport processes is not modeled, with it presumed to be nonlimiting. Accurate measurements of the kinetics for ATP binding and hydrolysis have yet to be determined and hence, at present, this assumption must stand. Indeed the available data suggests that transport data can be fitted without modeling this component, suggesting that it has a minimal impact on the overall flux.

#### 4.3.3 Measurement of Passive Diffusion

Derivation of a logP or logD for the passive permeability of a chemical is relatively trivial and can be carried out in a simple cell-free system (Wils et al. 1994). However, it should be noted that this simplistic equation presumes that there is a nonlimiting surface area through which the chemical can diffuse, and that the permeability of a lipid bilayer remains constant. If either of these assumptions is breached then it is necessary to include further terms to account for these parameters, derived from Fick's first and second laws of diffusion (Fick 1855, (4.5) and (4.6)).

$$J = -D \frac{\delta \phi}{\delta x} , \quad (4.5)$$

$$\frac{\delta \phi}{\delta t} = D \frac{\delta^2 \phi}{\delta x^2} , \quad (4.6)$$

where

$J$  is the the flux through the membrane ( $\text{M} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$ ),  $D$  is the diffusion coefficient ( $\text{cm}^2 \cdot \text{s}^{-1}$ ),  $\phi$  is the concentration (M), and  $x$  is the position (cm).

For the purpose of general modeling of chemical flux, Fick's laws may be expressed as shown in (4.7), which can be rewritten as the ordinary differential equation shown in (4.8).

$$J = -P \cdot \Delta \phi , \quad (4.7)$$

$$\frac{\delta A_{\text{out}}}{\delta t} = -\frac{P \cdot A}{V_{\text{in}}(A_{\text{out}} - A_{\text{in}})} , \quad (4.8)$$

where

$P$  is the permeability of the membrane ( $\text{cm} \cdot \text{s}^{-1}$ ),  $A$  is the surface area of the membrane ( $\text{cm}^2$ ),  $V_{\text{in}}$  is the volume of acceptor ( $\text{cm}^3$ ),  $A_{\text{in/out}}$  is the concentration of drug in donor and acceptor compartments (M). It should be noted that the units for the term  $PA/V$  cancel out to produce  $\text{s}^{-1}$ , which is consistent with this term representing the rate constant for passive diffusion.

*Correction for Multiple Sampling.* One potential issue with any system used to produce kinetic measurements is the requirement for multiple measurements to be taken over time. This may result in a decrease in the volume of, for example, the acceptor compartment as samples are taken, and it can be envisaged that this could become significant if extended time or large sampling volumes are used. Two possible solutions exist for this problem: First, the use of chemicals such as carboxydichlorofluorescein, which can be measured due to their fluorescence, and hence do not require a reduction in the sampling volume (Howe et al. 2009). Second, Tran et al. (2004) suggested the inclusion of an additional term into (4.8) to allow automatic correction for multiple sampling.

*Measurement of Passive Diffusion at Different Temperatures.* Previously, it has been assumed that passive diffusion was a temperature-independent phenomenon. This assumption allows measurement of passive diffusion parameters to be undertaken at  $4^\circ\text{C}$ , when active transport processes do not occur, thus providing an easy means to separate passive and active transport components for any given chemical. However, it is becoming increasingly clear that membrane fluidity, and hence permeability, are altered by temperature (Ulrich et al. 2007), and hence measurement of passive diffusion at  $4^\circ\text{C}$  can potentially be artefactual. For example, Poirier et al. (2008) demonstrated that the apparent permeability (Papp) of chemicals could be considerably lower at  $4^\circ\text{C}$  compared to  $37^\circ\text{C}$ , with the Papp for fexofenadine being 16-fold lower when determined at  $4^\circ\text{C}$  compared to  $37^\circ\text{C}$ .

If it is necessary to assess passive diffusion at  $37^\circ\text{C}$  in order to achieve accurate kinetic values then it is important to be able to distinguish active and passive components of transport, both of which can occur at this temperature. One approach is to undertake the assessment of passive diffusion in artificial membranes, such as the parallel artificial membrane permeability assay (PAMPA), or through the use of lipid vesicles, both of which do not contain any active transport proteins. The advantage of this system is the specificity of the measurement gained, but this comes at the cost of the use of a more artificial system that may breach some of the general assumptions outlined in Sect. 4.3.2. An alternative approach is to undertake assays in cell lines, thus increasing the closeness to *in vivo*, but with the use of chemical inhibitors to prevent active transport. The obvious disadvantage of such a system is that one must know something about the active transport of a chemical before one knows which inhibitors to use. Fortunately, there are now a number of broad specificity inhibitors, such as quinidine and verapamil, which allow one to inhibit large families of transport proteins at one time (Tan et al. 2000).

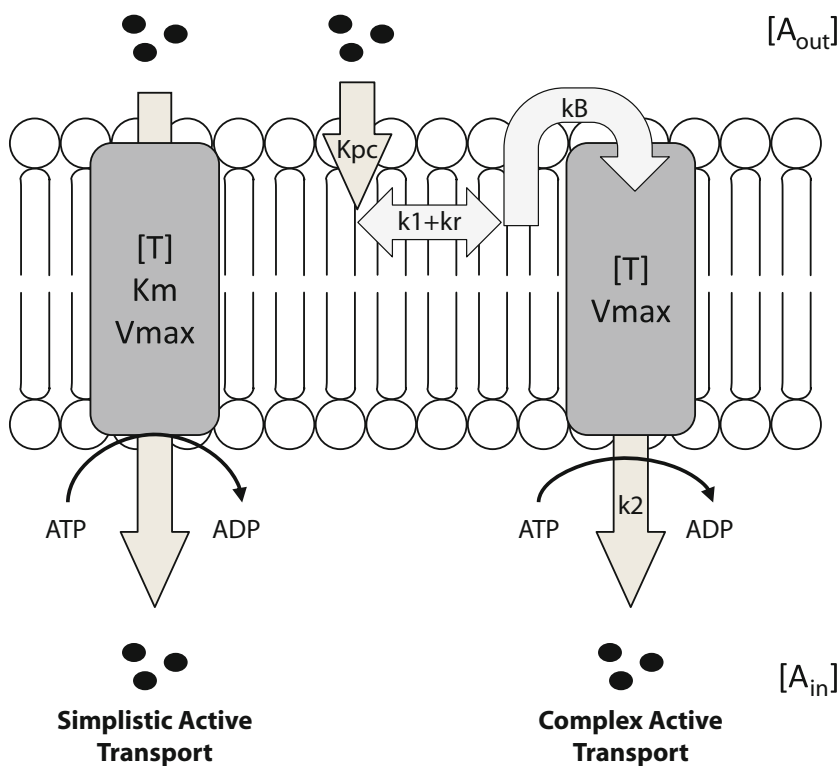
### 4.3.4 Measurement of Active Transport

The measurement of transporter-mediated movement across biological membranes is potentially much more complicated than the measurement of the passive diffusion component. Numerous test systems exist to examine active transport processes in cells. These can range from whole cells assays undertaken in Transwells™, which measure the total active transport kinetics rather than any single transport protein (Buesen et al. 2002), through over-expressing cell lines (Hopper-Borge et al. 2004) to (inside out) lipid-vesicles (Glavinas et al. 2008), the latter two of which allow the measurement of kinetics for a single transport protein. An advantage of membrane vesicles is that they can be engineered so that they are either orientated “normally” or “inside out”, meaning that both influx and efflux proteins can be studied with relative ease (Glavinas et al. 2008). Such an approach negates the need to preload cells with chemical for the study of efflux transport, which is an additional complication that can require optimization.

Regardless of the system of study used, it is usual to fit chemical kinetics to the Michaelis–Menten (4.2), which has certain limitations. The Michaelis–Menten analysis was derived from the analysis of soluble enzyme kinetics and hence may not be suitable for examining the kinetics of membrane-bound transporters. The largest difference between these two scenarios is the location of chemical prior to interaction with the enzyme or transporter. In the case of soluble enzymes the chemical will generally be in the aqueous phase (i.e., cytoplasm) and hence can directly interact with the enzyme. In the case of transporters, however, the chemical may be present within either the cytoplasm, as occurs with glucose transport, for example, or embedded within the lipid membrane if it is sufficiently lipophilic, as is the case for many drug molecules. In both cases diffusion through the cellular environment will impact upon the rate of association with transporters, but it is logical to hypothesize that this would be of greater impact for diffusion within the membrane compared to the cytoplasm. One solution is to include a scaling factor, similar to the Hill function within the transport equation (4.3). This scaling factor does not necessarily represent multiple binding sites or cooperativity within an individual protein, as is the case for soluble enzymes, but represents the multiple steps that interact to provide the total measured kinetics; this scaling factor is often referred to as the  $\beta$ -coefficient. This solution works well, but has the disadvantage that the all-encompassing scaling factor hides many of the subtleties involved in transport, which may be critical in understanding the true biological mechanisms occurring within the transport process.

Bentz et al. (2005) have attempted to deconvolute the individual steps in drug transport, measuring the individual kinetic parameters shown in Fig. 4.3 (Tran et al. 2005), and fitting these into a model of chemical transport.

As can be seen from Fig. 4.3, rate constants were derived for each of the stages involved in drug transport. These included association of the chemical into the lipid membrane from the aqueous solution, which was chemical specific and highly variable ( $K_{pc}$ ); diffusion through the lipid membrane, allowing association with the transporter, which appears to be both fast and nonchemical specific ( $k_1/k_r$ ); binding



**Fig. 4.3** Simple and complex determinations of active transport kinetics. In the simple mode, the increase in intracellular drug concentration  $[A_{in}]$  is determined by the external drug concentration  $[A_{out}]$ , transporter concentration  $[T]$ , and efficiency of the transporter toward the drug ( $K_m$  and  $V_{max}$ ). In comparison, the complex model takes into account these three factors plus association rate of the drug into the lipid membrane ( $K_{pc}$ ), diffusion rate within the membrane ( $k_1$  and  $k_r$ ), association rate of the drug with the transporter ( $k_B$ ) and disassociation rate of the drug from the transporter on the far side of the membrane ( $k_2$ )

of the chemical to the transport channel of the transport protein, which is chemical specific ( $K_B$ ); and the efflux of the chemical to the other side of the membrane, again chemical specific ( $k_2$ ). So, does this increased detail within the transport model translate to improved biological knowledge and predictability? Bentz et al. (2005) examined this by simulating the transport process using physiologically relevant parameter ranges. They were able to demonstrate that modeling of  $V_{max}$  was relatively robust, with only a two- to three-fold variance between the modeled and experimental  $V_{max}$  values. It should be noted, however, that this conclusion was only valid when the efflux velocity measured was significantly lower than the  $[A_{out}]$ , meaning that  $V/[A_{out}]$  tended toward zero. Such a scenario was necessary to ensure that the passive diffusion component of chemical transport did not become too large and equated to a transporter occupancy rate of over 80%. This is of interest because such a rate of occupancy may not always be achieved under experimental

conditions. First, systems using over expression plasmids for the transport protein under examination may have many more copies of the protein expressed than is seen in vivo and thus achieving the  $[A_{\text{out}}]$  may be difficult. Second, as stated in the general assumptions for modeling (Sect. 4.3.2), as chemical concentration increases it is probable that the CMC will be breached and that the dynamics of chemical interactions with the transporter altered, effectively increasing the input of passive diffusion into the model. As will be discussed in Sect. 4.4.1, one solution to this issue is to create separate terms for both the passive and active components of the transport, which can then be modulated independently.

Whereas  $V_{\text{max}}$  could be reasonably well modeled over the physiological range using simple equations such as (4.2) and (4.3), the derivation of the overall  $K_m$  was not as simple. This is perhaps not surprising given that several of the rate constants impact upon the  $K_m$ . Bentz et al. (2005) demonstrated that derivation of biological  $K_m$  values from models was subject to considerable inaccuracy, which increased as  $K_m$  did. This latter point is especially important when considering the transport of drug-like molecules, as they are routinely transported by promiscuous transport proteins with high, micromolar,  $K_m$  values. Bentz et al. (2005) suggested that the biological  $K_m$  could be modeled from the elementary rate constants as shown in (4.9), recommending that elementary rate constants provide a better estimate of  $K_m$  than fitting with the simple Michaelis–Menten equation.

$$K_{m\text{biol}} = \frac{k_2 + k_r}{k_1 + K_{\text{pc}}}, \quad (4.9)$$

where

$K_{m\text{biol}}$  is the “biologically relevant”  $K_m$  (M);  $k_2$  is the disassociation rate constant for drug from transporter;  $k_r$  and  $k_1$  are the diffusion rate constants for drug within the membrane, often equal; and  $K_{\text{pc}}$  is the association rate constant for drug and lipid membrane.

#### 4.4 The Integration of Drug Disposition and Drug Fate into a Predictive Model of the Life Cycle of a Drug in the Body

The understanding of the kinetics of drug entry into cells is an area that has received much attention in the past few years, with an ever increasing understanding of both the importance and complexity of this process becoming clear (Ayrton and Morgan 2001). However, this biological understanding is really only the first stepping stone to understanding larger biological questions, such as the life cycle of chemicals within the body. Any chemical, whether it is endogenous or exogenous in origin, is subject to the processes of ADME, which together determine the longevity and action of the chemical within the body. As described in Sect. 4.2, the science of

pharmacokinetics has developed specifically to allow these processes to be modeled, with the ultimate aim of allowing accurate extrapolation and prediction of biological responses in humans, as detailed in Sect. 4.1.3. One approach for the incorporation of membrane transport into larger models is the integration of multiple processes into single formula to cover all aspects of transport. However, as the amount of information on the mechanics of transports becomes clear then these equations become more complicated and cumbersome.

Whereas the integration of multiple processes into single equations is valid, as the complexity of the derived equations increases they rapidly become impenetrable to all but experts, thus diminishing their utility by the general scientific community. An alternate approach is the use of modeling software, which is similar to pathway mapping software originally designed to understand DNA microarray transcript data, providing a user-friendly front-end for complex simulations. Programs such as CellDesigner<sup>6</sup> (Funahashi et al. 2003) provide an easy to use graphical front-end that allows nonspecialist entry of network maps in Systems Biology Graphical Notation<sup>7</sup> (SBGN), an established format that is widely utilized. In addition, whereas the model is created by a simple point-and-click interface it is encoded by the Systems Biology Markup Language<sup>8</sup> (SBML; Hucka et al. 2003), a universal XML format, which allows the model to be easily transferred between modeling software. To provide an example of the utility of such an approach we will examine the modeling of multiple drug resistance phenotype in the treatment of breast cancer.

#### ***4.4.1 Multiple Drug Resistance Phenotype in Cancer Treatment***

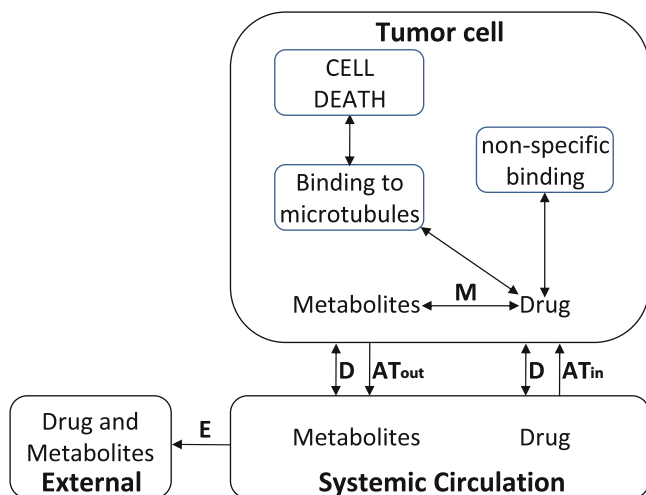
More than one million women are diagnosed with breast cancer every year, which represents approximately one quarter of all new cancers in women: Such a rate of diagnosis represents a lifetime risk of developing breast cancer of one in eight for women born in the USA, and one in nine for women in the United Kingdom (Coley 2008). Fortunately, there are a number of established therapies for metastatic breast cancer (MBC), ranging from endocrine-based therapies for hormone-receptor positive tumors, through anthracyclines and taxanes to the recent development of novel biologics such as trastuzumab (Herceptin). Despite this range of therapeutics, the response rate to first-line chemotherapies such as anthracyclines and taxanes is sub-optimal, being reported as between 30 and 70%, falling to 20–30% for subsequent treatments with a median duration of response of 6 months. One area that often limits chronic therapeutic treatment, including anti-cancer chemotherapy, is the development of multiple drug resistance (MDR) phenotype, and it has been estimated that MDR is involved in over 90% of treatment failures for MBC.

---

<sup>6</sup> <http://www.celldesigner.org>.

<sup>7</sup> <http://sbgn.org/Main.Page>.

<sup>8</sup> <http://sbml.org/Main.Page>.



**Fig. 4.4** *Mode of action of Taxane-family anticancer drugs.* Taxanes can be taken into cells by either passive diffusion ( $D$ ) or active transport ( $AT_{in}$ ). Once in the cell they can either bind non-specifically, or specifically to microtubules. In the latter case this elicits a chain of events that leads to cell death, killing the tumor cell. Taxanes can also be metabolized in the cell by several different enzymes, and these metabolites removed from the cells by passive diffusion ( $D$ ) or active transport ( $AT_{out}$ ). Finally, these products can be removed from the body by excretion ( $E$ )

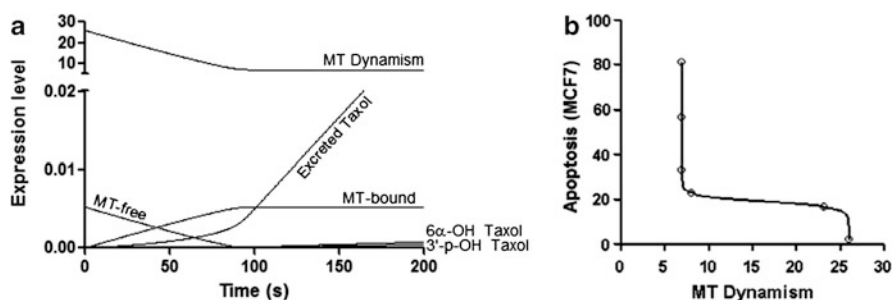
The taxane class of anticancer drugs are a front-line treatment for metastatic breast cancer, with paclitaxel (Taxol) being the class leader (McGrogan et al. 2008). Their mode of action involves high affinity binding to microtubules, cellular structures oblige for biological functioning (Fig. 4.4). Taxol binding prevent microtubules from growing and shortening, which is essential for their functioning, and this causes the cell to undergo a series of steps that results, ultimately, in cell death (Xiao et al. 2006). Prolonged treatment with taxol often results in the production of MDR, preventing the drug effectively killing cells. Several mechanisms have been identified that will contribute toward the development of MDR; increased metabolism, increased export, altered interaction with target proteins and altered biological response to that interaction. However, it is not clear how each of these factors input into the development of MDR, and hence what is the best strategy to prevent its development: Such a problem is ideally suited to examination using a systems modeling approach.

Based on the cartoon shown in Fig. 4.4, a comprehensive model of the network of interactions of taxol with tumor cells was created in CellDesigner. All interactions within the network were derived from published literature, producing a model with 123 species and 72 reactions. Underlying the graphical front-end, each reaction can be associated with a kinetic law, producing a set of ordinary differential equations (ODE) to describe the network. By dividing the interaction network into a series of ODEs it is possible to build the network model from a series of interconnected modules; in the example shown, for example, individual modules might

represent cellular influx, metabolism, sequestration, efflux, binding to microtubules and triggering of programmed cell death. Such an approach is ideal as it reduces the knowledge base required, both biochemical and mathematical, making the approach available to a larger number of scientists. In addition, as each reaction is defined by simple ODEs then as new information becomes available it is easy to replace one, or a number, or reactions with the new details, thus improving the model.

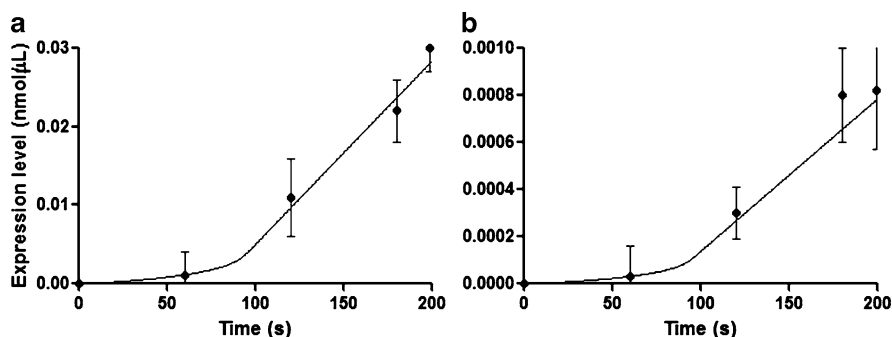
Once such models have been generated, the next step is to undertake simulations of the entire network. Such analysis can be undertaken directly in CellDesigner, but is limited to relatively simple time simulations and parameter scans. A more flexible alternative is to export the model into more powerful analysis software such as COPASI<sup>9</sup> (Hoops et al. 2006), and this is where the use of the universal SBML format makes such transitions simple. Once in COPASI further analysis options are available, including steady-state analysis, parameter estimation, network optimization, and sensitivity analysis; in addition, both deterministic and stochastic modeling is available, an important requirement for examination of single cell events. More experienced modelers may wish to program directly into COPASI, which uses a tabular format to define reactions, species, etc.; however, the lack of a graphic representing the overall network can be confusing for inexperienced modelers. Therefore, it is often easier to build outline models in programs such as CellDesigner due to their simple model entry system, before porting the model to a comprehensive simulation tool such as COPASI.

Having transferred the taxol model into COPASI, a first analysis over a time series shows that, after taxol addition, there is a rapid decrease in the amount of free microtubules and concomitant increase in taxol-bound microtubules (Fig. 4.5a). The net effect of this is a decrease in microtubule dynamism, and as can be seen from panel b, microtubule dynamism is directly related to the rate of apoptosis,



**Fig. 4.5** Time series simulation of Taxol in MCF7 breast cancer cells. A model of the interactions of taxol in MCF7 breast cancer cells was created in CellDesigner, transferred to COPASI and the network simulate over 200 s, demonstrating a decrease in microtubule dynamism (MT; Panel a). Decreases in microtubule dynamism can be linked to the rate of programmed cell death, apoptosis (Panel b)

<sup>9</sup> <http://www.copasi.org>.



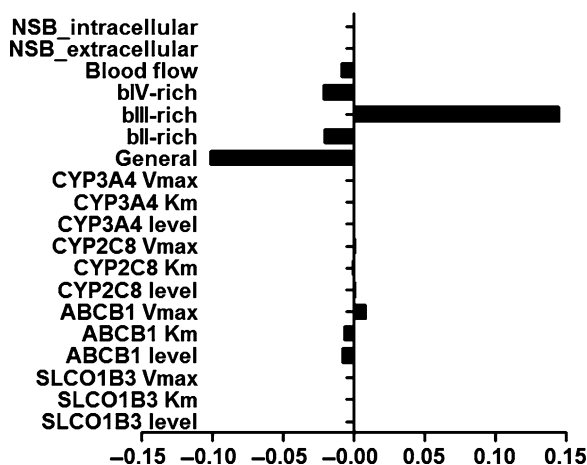
**Fig. 4.6** Time series simulation of Taxol in MCF7 breast cancer cells. A model of the interactions of taxol in MCF7 breast cancer cells was created in CellDesigner, transferred to COPASI and the network simulated over 200 s (line). These values were compared to measurements made in vitro (closed circles) for excreted taxol (Panel a) and production of the 6 $\alpha$ -OH taxol metabolite (Panel b)

programmed cell death, in breast cancer cells (MCF7 cell line; Fig. 4.5b). However, this desired, pharmacological action is offset by excretion of taxol from the cell by drug transporters (excreted taxol), and the metabolism of taxol (6 $\alpha$ -OH and 3'-p-OH taxol).

Having created the model, it is important to show that the simulated species levels closely match those seen in the biological scenario. Whereas it is not possible to easily measure all of the parameters, it is important to show that as many as possible correlate well, as demonstrated in Fig. 4.6.

Following demonstration of correlation between the in silico and in vitro model systems it is now possible to use the in silico model to examine the underlying design principles for the network. As an initial step in this process a sensitivity analysis will reveal those nodes within the network most likely to impact upon MT dynamism. As depression of MT dynamism is central to the anticancer effects of taxol, any node that significantly impacts upon this is a potential mechanism for the development of multiple drug resistance phenotype. As can be seen from Fig. 4.7, such an analysis highlights three factors (microtubule subunit composition, blood flow to tissue and ABCB1 transporter expression) as being the probable major drivers for development of multiple drug resistance phenotype. One of these species is the drug transporter ABCB1, demonstrating the importance of membrane transport in determining drug disposition and actions.

Having gained this information it is now possible to use molecular techniques to specifically target these species for intense examination within the in vitro system. For example, cell lines can be genetically modified to allow alteration of the species in question, varying their expression in line with that observed in vivo. These modified cell lines can then be used to confirm the predictions of the in silico model, as well as acting as test systems to examine how the development of MDR could be prevented in the clinic. In addition, other biological questions can be answered



**Fig. 4.7** *Sensitivity analysis.* A model of the interactions of taxol in MCF7 breast cancer cells was created in CellDesigner, and then transferred to COPASI for sensitivity analysis. Sensitivity of species within the network were determined against their ability to alter the impact of taxol on microtubule dynamism

surrounding drug treatment of cancer, such as the extrapolation of response to Taxol in normal and tumor tissue, where protein levels will be different; to extrapolate between preclinical and clinical scenarios; finally, novel or altered therapeutic treatments can be simulated to assess their potential utility.

## 4.5 Summary

The cell membrane is an essential barrier for life. It acts not only to contain the cell contents but also as a barrier to prevent free access of chemicals to the cell interior. As such, understanding the dynamics of chemical transport across biological membranes is vital to fully understand how the body will respond to chemical exposure. Recent work has begun to characterize this, perhaps surprisingly complex, mechanism, and is increasing our understanding of the determinants for chemical access to cell interiors. The integration of these transport mechanisms into larger models of general cellular response to chemical exposure is currently underway, producing more refined models of the life cycle and biological impact of chemical exposure. Such approaches will be vital for the extrapolation of data between different biological scenarios, such as understanding the differential response of normal and tumor tissue and the development of multiple drug resistance phenotype, and lead to improved treatment schedules with increased success rates.

## References

- P. Acharya, T. T. Tran, J. W. Polli, A. Ayrton, H. Ellens, and J. Bentz. P-glycoprotein (P-gp) expressed in a confluent monolayer of hMDR1-MDCKII cells has more than one efflux pathway with cooperative binding sites. *Biochemistry*, 45(51):15505–15519, 2006
- S. Aouabdi, G. G. Gibson, and N. Plant. Transcriptional regulation of PXR: Identification of a PPRE within the proximal promoter responsible for fibrate-mediated transcriptional activation of PXR. *Drug Metabolism and Disposition*, 34(1):138–144, 2006
- A. Ayrton and P. Morgan. Role of transport proteins in drug absorption, distribution and excretion. *Xenobiotica*, 31(8–9):469–497, 2001
- J. Bentz, T. T. Tran, J. W. Polli, A. Ayrton, and H. Ellens. The steady-state Michaelis-Menten analysis of P-glycoprotein mediated transport through a confluent cell monolayer cannot predict the correct Michaelis constant  $K_m$ . *Pharmaceutical Research*, 22(10):1667–1677, 2005
- J. G. M. Bessems and N. P. E. Vermeulen. Paracetamol (acetaminophen)-induced toxicity: Molecular and biochemical mechanism, analogues and protective approaches. *Critical Reviews in Toxicology*, 31(1):55–138, 2001
- R. Buesen, M. Mock, A. Seidel, J. Jacob, and A. Lampen. Interaction between metabolism and transport of benzo[a]pyrene and its metabolites in enterocytes. *Toxicology and Applied Pharmacology*, 183(3):168–178, 2002
- H. M. Coley. Mechanisms and strategies to overcome chemotherapy resistance in metastatic breast cancer. *Cancer Treatment Reviews*, 34(4):378–390, 2008
- A. Fick. Ueber diffusion. *Annalen der Physik*, 170:59–86, 1855
- A. Funahashi, N. Tanimura, M. Morohashi, and H. Kitano. CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1:159–162, 2003
- G. G. Gibson and P. Skett. *Introduction to drug metabolism*. Nelson Thornes, Cheltenham, 3rd edition, 2001
- H. Glavinas, D. Mehn, M. Jani, B. Oosterhuis, K. Heredi-Szabo, and P. Krajcsi. Utilization of membrane vesicle preparations to study drug-ABC transporter interactions. *Expert Opinion in Drug Metabolism and Toxicology*, 4(6):721–732, 2008
- S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI – a COMplex PATHway SIMulator. *Bioinformatics*, 22:3067–3074, 2006
- E. Hopper-Borge, Z. S. Chen, I. Shchaveleva, M. G. Belinsky, and G. D. Kruh. Analysis of the drug resistance profile of multidrug resistance protein 7 (ABCC10): Resistance to docetaxel. *Cancer Research*, 64(14):4927–4930, 2004
- K. Howe, G. G. Gibson, T. Coleman, and N. Plant. In silico and in vitro modelling of hepatocyte drug transport processes: Importance of ABCC2 expression levels in the disposition of carboxydichlorofluorescein. *Drug Metabolism and Disposition*, 37(2): 391–399, 2009
- M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003
- S. H. Kim, F. R. Miller, L. Tait, J. Zheng, and R. F. Novak. Proteomic and phosphoproteomic alterations in benign, premalignant and tumor human breast epithelial cells and xenograft lesions: Biomarkers of progression. *International Journal of Cancer*, 124(12):2813–2828, 2009
- H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004
- H. Lennernas. Intestinal permeability and its relevance for absorption and elimination. *Xenobiotica*, 37(10–11):1015–1051, 2007

- B. T. McGrogan, B. Gilmartin, D. N. Camey, and A. McCann. Taxanes, microtubules and chemoresistant breast cancer. *Biochimica Et Biophysica Acta-Reviews on Cancer*, 1785(2):96–132, 2008
- K. Palm, K. Luthman, J. Ros, J. Grasjo, and P. Artursson. Effect of molecular charge on intestinal epithelial drug transport: pH-dependent transport of cationic drugs. *Journal of Pharmacology and Experimental Therapeutics*, 291(2):435–443, 1999
- J.-M. Pascussi, S. Gerbal-Chaloin, J.-M. Fabre, P. Maurel, and M.-J. Vilarem. Dexamethasone enhances constitutive androstane receptor expression in human hepatocytes: Consequences on cytochrome P450 gene regulation. *Molecular Pharmacology*, 58(6):1441–1450, 2000a
- J.-M. Pascussi, L. Drocourt, J.-M. Fabre, P. Maurel, and M.-J. Vilarem. Dexamethasone induces pregnane X receptor and retinoid X receptor- $\alpha$  expression in human hepatocytes: Synergistic increase of CYP3A4 induction by pregnane X receptor. *Molecular Pharmacology*, 58:361–372, 2000b
- N. Perriere, S. Yousif, S. Cazaubon, N. Chaverot, F. Bourasset, S. Cisternino, X. Decleves, S. Hori, T. Terasaki, M. Deli, J. M. Scherrmann, J. Temsamani, F. Roux, and P. O. Couraud. A functional in vitro model of rat blood-brain barrier for molecular analysis of efflux transporters. *Brain Research*, 1150:1–13, 2007
- N. Plant. *Molecular toxicology*. Advanced Text. BIOS, London, 2003
- N. Plant. Interaction networks: Coordinating responses to xenobiotic exposure. *Toxicology*, 202: 21–32, 2004
- N. Plant. The human cytochrome P450 3A sub-family: Transcriptional regulation, inter-individual variation and interaction networks. *Biochimica Biophysica Acta*, 1770(3):478–488, 2007
- A. Poirier, T. Lave, R. Portmann, M. E. Brun, F. Senner, M. Kansy, H. P. Grimm, and C. Funk. Design, data analysis, and simulation of in vitro drug transport kinetic experiments using a mechanistic in vitro model. *Drug Metabolism and Disposition*, 36(12):2434–2444, 2008
- S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences United States of America*, 98(26):15149–15154, 2001
- G. Rechekmmer. Transport of weak electrolytes. In M. Field and R. A. Frizzel, editors, *Handbook of physiology, section 6: The gastrointestinal system*, pp 371–388. American Physiological Society, Bethesda, 1991
- R. A. Roberts, D. W. Nebert, J. A. Hickman, J. H. Richburg, and T. L. Goldsworthy. Perturbation of the mitosis/apoptosis balance: A fundamental mechanism in toxicology. *Fundamental and Applied Toxicology*, 38(2):107–115, 1997
- A. Rostami-Hodjegan and G. T. Tucker. Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nature Reviews Drug Discovery*, 6(2):140–148, 2007
- G. A. Sawada, C. L. Barsuhn, B. S. Lutzke, M. E. Houghton, G. E. Padbury, N. F. H. Ho, and T. J. Raub. Increased lipophilicity and subsequent cell partitioning decrease passive transcellular diffusion of novel, highly lipophilic antioxidants. *Journal of Pharmacology and Experimental Therapeutics*, 288(3):1317–1326, 1999
- G. Shinar, R. Milo, M. R. Martinez, and U. Alon. Input-output robustness in simple bacterial signaling systems. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19931–19935, 2007
- M. D. Slack, E. D. Martinez, L. F. Wu, and S. J. Altschuler. Characterizing heterogeneous cellular responses to perturbations. *Proceedings of the National Academy of Sciences United States of America*, 105(49):19306–19311, 2008
- D. Smirlis, R. Muangmoonchai, M. Edwards, I. R. Phillips, and E. A. Shephard. Orphan receptor promiscuity in the induction of cytochromes P450 by xenobiotics. *Journal of Biological Chemistry*, 276(16):12822–12826, 2001
- C. Sotiriou and L. Pusztai. Gene-expression signatures in breast cancer. *New England Journal of Medicine*, 360(8):790–800, 2009

- B. Tan, D. Piwnica-Worms, and L. Ratner. Multidrug resistance transporters and modulation. *Current Opinion in Oncology*, 12(5):450–458, 2000
- T. N. Tozer and M. Rowland. *Introduction to pharmacokinetics and pharmacodynamics: The quantitative basis of drug therapy*. Lippincott Williams & Wilkins, Philadelphia, 2006
- T. T. Tran, A. Mittal, T. Gales, B. Maleeff, T. Aldinger, J. W. Polli, A. Ayrton, H. Ellens, and J. Bentz. Exact kinetic analysis of passive transport across a polarized confluent MDCK cell monolayer modeled as a single barrier. *Journal of Pharmaceutical Sciences*, 93(8):2108–2123, 2004
- T. T. Tran, A. Mittal, T. Aldinger, J. W. Polli, A. Ayrton, H. Ellens, and J. Bentz. The elementary mass action rate constants of P-gp transport for a confluent monolayer of MDCKII-hMDR1 cells. *Biophysical Journal*, 88(1):715–738, 2005
- N. P. Ulrih, U. Adamljc, M. Nemec, and M. Sentjurc. Temperature- and pH-induced structural changes in the membrane of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *Journal of Membrane Biology*, 219(1–3):1–8, 2007
- D. J. Villeneuve, S. L. Hembruff, Z. Veitch, M. Cecchetto, W. A. Dew, and A. M. Parissenti. cDNA microarray analysis of isogenic paclitaxel- and doxorubicin-resistant breast tumor cell lines reveals distinct drug-specific genetic signatures of resistance. *Breast Cancer Research and Treatment*, 96(1):17–39, 2006
- M. L. H. Vlaming, J. S. Lagas, and A. H. Schinkel. Physiological and pharmacological roles of ABCG2 (BCRP): Recent findings in Abcg2 knockout mice. *Advanced Drug Delivery Reviews*, 61(1):14–25, 2009
- U. K. Walle and T. Walle. Taxol transport by human intestinal epithelial Caco-2 cells. *Drug Metabolism and Disposition*, 26(4):343–346, 1998
- R. E. Watkins, G. B. Wisely, L. B. Moore, J. L. Collins, M. H. Lambert, S. P. Williams, T. M. Willson, S. A. Kliewer, and M. R. Redinbo. The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science*, 292(5525):2329–2333, 2001
- P. Wils, A. Warnery, V. Phungba, S. Legrain, and D. Scherman. High lipophilicity decreases drug transport across intestinal epithelial-cells. *Journal of Pharmacology and Experimental Therapeutics*, 269(2):654–658, 1994
- H. Xiao, P. Verdier-Pinard, N. Fernandez-Fuentes, B. Burd, R. Angeletti, A. Fiser, S. B. Horwitz, and G. A. Orr. Insights into the mechanism of microtubule stabilization by Taxol. *Proceedings of the National Academy of Sciences of the United States of America*, 103(27):10166–10173, 2006
- K. A. Youdim, A. Avdeef, and N. J. Abbott. In vitro trans-monolayer permeability calculations: Often forgotten assumptions. *Drug Discovery Today*, 8(21):997–1003, 2003
- Y. Zhou, E. Hopper-Borge, T. Shen, X. C. Huang, Z. Shi, Y. H. Kuang, T. Furukawa, S. Akiyama, X. X. Peng, C. R. Ashby, X. Chen, G. D. Kruh, and Z. S. Chen. Cepharanthine is a potent reversal agent for MRP7(ABCC10)-mediated multidrug resistance. *Biochemical Pharmacology*, 77(6):993–1001, 2009

# Chapter 5

## Systems Biology of Tuberculosis: Insights for Drug Discovery

Karthik Raman and Nagasuma Chandra

### 5.1 Introduction

It is estimated that about two billion people, equalling one-third of the world's total population, are infected with *Mycobacterium tuberculosis* (*Mtb*) (World Health Organisation 2008). There are nearly two million deaths every year, translating to about four deaths a minute. Tuberculosis (TB) is also the leading killer among HIV-infected people with weakened immune systems. An additional problem we are confronted with in the recent years is the emergence of drug resistant varieties of TB. About 500,000 new multi-drug resistant TB (MDR-TB) cases are estimated to occur every year (World Health Organisation 2008).

More than 20 drugs and the Bacillus–Calmette–Guerin (BCG) vaccine are available for the treatment of TB. The existing drugs, although of immense value, have several shortcomings, the most important of them being the emergence of drug resistance, rendering even the front-line drugs inactive. In addition, drugs such as rifampicin have high levels of adverse effects making them prone for patient incompliance. Adding to these problems are the vicious interactions between the human immunodeficiency virus and TB, which lead to further challenges for anti-tubercular drug discovery (Nunn et al. 2005). For example, protease inhibitors have been shown to be incompatible with rifampicin-containing anti-TB regimens (Bonora and Di Perri 2008). The existence of several challenges in tackling TB necessitates the application of newer techniques to study and understand tubercular infection, as well as generate methods to counter it.

The genomics and the post-genomics eras, with the parallel advances in high-throughput experimental methods and screening techniques to analyse whole genomes and proteomes, are witnessing an explosion in the types and amount of information available, not only with respect to the genome sequences and protein

---

K. Raman (✉)

Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India

e-mail: [k.raman@bioc.unizh.ch](mailto:k.raman@bioc.unizh.ch)

structures but also with respect to gene-expression, regulation and protein–protein interactions. The availability of such information in publicly accessible databases and the advances in both computing power and computational methods for data mining and modeling have led to the emergence of several *in silico* approaches to systematically address important questions in biology, with an obvious impact on drug discovery (Apic et al. 2005; Claus and Underwood 2002). Systems-level approaches to drug discovery aid at multiple stages in the drug discovery pipeline, particularly in target identification and in identifying the molecular basis of disease for rational drug discovery.

Drug discovery in the past has relied heavily on animal models and *in vivo* studies. *In vitro* and biochemical studies have served mainly to back up the findings and provide mechanistic explanations where possible. It is obvious from these that the need for considering the system as a whole has always been well recognised. It can be argued that using a mouse or any typical animal model is also a systems approach. While that is true in some sense, such approaches deviate significantly from the current systems practices, since the former is more a “black box”, which enables a “readout” that is a systems output, but does not tell us why or how such an output results. The current practices, on the other hand, attempts to reconstruct the system brick by brick and hence facilitates an understanding of why and how an event takes place, automatically leading to “what if” type of questions, enabling predictions. The current approaches also depart from the “spherical cow” abstractions that have often characterised mathematical modeling (Doyle 2001), by virtue of starting reconstructions from thousands of fundamental building blocks leading to “realistic” modeling. Availability of “omics” scale experimental data on various fronts such as genome sequencing, transcriptional profiles, proteins expressed, lipids, glycans and the hundreds of metabolites that are interconverted by the molecular metabolic machinery indeed facilitate such realistic modeling.

This chapter focuses on some of the recent advances in the understanding of *Mtb*, from a systems biology perspective and the potential of systems-level analyses to generate more useful drug targets and a better understanding of the disease and the pathogen itself. The chapter also discusses possibilities of application of systems approaches in understanding important issues that arise in drug discovery, such as interaction between the drug, the target and the system as a whole, possible side effects and causes of drug toxicity. The complete knowledge of metabolic reactions in an organism helps to analyse all possible interactions between the drug and the system and also helps to narrow down possible causes for adverse effects and drug toxicity. Given the fact that cellular systems are extremely complex, a systematic analysis of all reactions taking place in a cell across various biochemical pathways is a challenging task. The following sections illustrate some of the approaches that have been taken to understand *Mtb* metabolism, protein–protein interactions, emergence of resistance to anti-TB drugs, as well as the complex interactions of *Mtb* with the host immune system.

## 5.2 Understanding *Mtb*: A Parts Catalogue

Systems are composed of individual elements or “parts” that interact in various ways. In general, the behaviour of a system is quite different from merely the sum of the interactions of its various parts, applicable even more so for complex biological systems. As Anderson put it as early as 1972 in his classic paper by the same title, “More is different” (Anderson 1972), it is not possible to reliably predict the behaviour of a complex system, despite a good knowledge of the fundamental laws governing the individual components. Systems biology emphasises the study of larger systems, in an attempt to better elucidate the complex web of interactions between various underlying components of biological systems.

Every single cell is made up of a bewildering variety of molecules, macromolecules and their complexes. In *Mtb*, its cell wall itself is an excellent example of complexity: it is made up of three polymers, arabinogalactan-mycolate (Crick et al. 2001), covalently linked with peptidoglycan and trehalose dimycolate, which provide a thick protective layer from general antibiotics and the human immune system (Takayama et al. 2005). Beneath this surface lies a metabolic network with about a thousand metabolites ranging from simple carbohydrates to complex lipids and long-chain fatty acids, facilitated and regulated by an array of proteins.

## 5.3 Assembling the Parts: Network Reconstruction

Systems biology, being a holistic approach to study biological systems in contrast to traditional reductionist approaches, involves the synthesis of models of the various “parts” discussed above into networks depicting metabolism, regulation, signalling and protein–protein interactions, by a process usually termed as “reconstruction”. Reconstruction involves the integration of disparate sources of data to create a representation of the chemical events underlying the different biological networks (Papin et al. 2005).

Table 5.1 gives a broad overview of the various components of the different biological networks, as well as the elements involved in their reconstruction and methods for simulation. Figure 5.1 presents a graphical view of the various levels of hierarchy at which *Mtb* can be studied, also indicating the computational methods that are generally used.

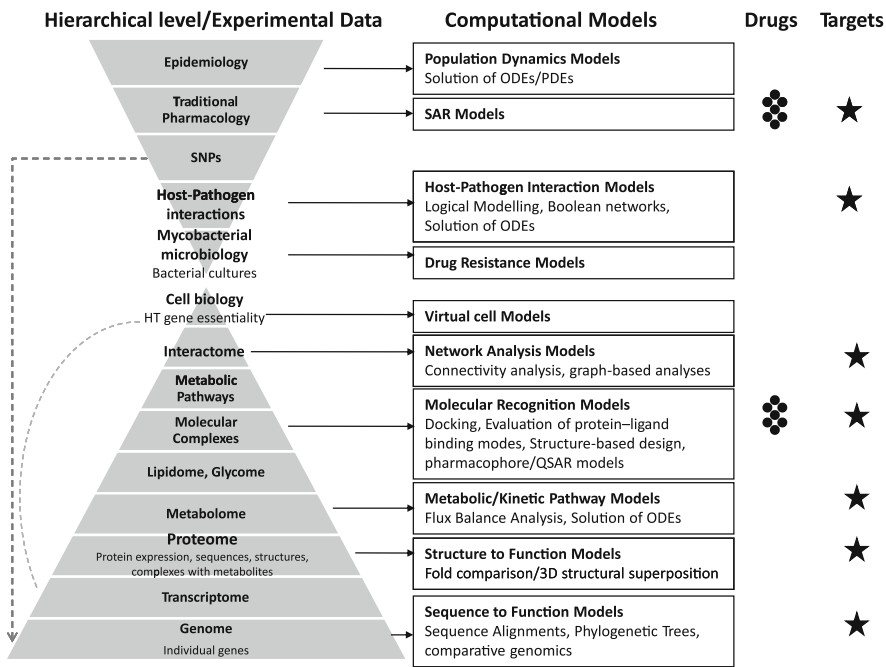
### 5.3.1 Annotation of Genomes

The sequencing of the entire genome by Cole and co-workers (Cole et al. 1998), a landmark in TB research, provided the first glimpse of the genomic constitution leading to deciphering the nearly 4,000 genes in it and their protein products. This finding has triggered significant downstream research in the area, many of

**Table 5.1** *Overview of the reconstruction of biological networks.* The major network types and their components, elements of their reconstruction as well as methods for their simulation are listed here

|                            | Metabolic networks                                                                                                                                | Signal transduction networks                                                                                               | Transcriptional regulatory networks                                                                                           | Protein–protein interaction networks                                                                                       |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Components                 | Metabolites<br>Proteins<br>(Enzymes)<br>Reactions                                                                                                 | Proteins<br>Ions<br><br>Metabolites                                                                                        | Operons<br>Regulons<br><br>Stimulons                                                                                          | Evolutionary/<br>functional/<br>structural<br>linkages<br>between<br>proteins                                              |
| Elements of reconstruction | Genomic data (annotations)<br>Stoichiometry<br>Gene–protein–reaction associations<br>Reaction rates<br>Cellular constraints<br>Kinetic parameters | Nodes<br>Modules<br>Motifs<br>Functional data<br>Protein–protein interactions<br>Reaction mechanisms<br>Kinetic parameters | Component data<br>Interaction data<br>Network state data<br>Reaction mechanisms<br>Kinetic parameters<br>Causal relationships | Phylogenetic profiling<br>Rosetta stone<br>Gene neighbourhood<br>Operon<br>In silico<br>two-hybrid<br>Experimental methods |
| Methods for simulation     | Stoichiometric analysis<br>Constraint-based methods<br>Interaction-based modeling (graphs)<br>Mechanistic modeling (differential equations)       | Boolean networks<br>Mechanistic modeling<br>Ensemble modeling                                                              | Mechanistic modeling<br>Stoichiometric analysis<br>Boolean networks                                                           | Topological analysis                                                                                                       |

them at the “omics” scale, such as proteomics, transcriptomics and metabolomics, through newer technologies. Comprehension of the large oceans of such data and translation to useful biological insights require an understanding of what the individual genes and proteins in the genome do. Genome annotation is in fact a very important and critical requirement to leverage benefit from large-scale data. Advances in bioinformatics have led to the development of several tool-kits, which have to a major extent evolved rapidly in the last two decades to meet the demand stemming from whole genome sequencing. Through bioinformatics analyses, a significant amount of the biology of the bacillus has been deciphered through the identification of genes and proteins involved in several functional modules such as core metabolic pathways, characteristic lipid metabolism, polyketide



**Fig. 5.1** Various levels of hierarchy at which *Mtb* can be modeled. The various levels of hierarchy at which *Mtb* can be modeled, and the experimental data that are available are illustrated. Models at many of these levels are useful for drug target identification and drug discovery. The lower pyramid illustrates the different levels of organisation in the cell. While metabolic pathways and the interactome are more commonly analysed in systems biology studies, the significant interplay between the levels necessitates the consideration of many other levels, such as the transcriptome and the metabolome, which present many insights into the complex web of interactions in a cell

and siderophore metabolism, insertion sequences, immunity and pathogenicity determinants (Cole et al. 1998; Camus et al. 2002). These annotations pave way for higher order reconstructions of modules in the genome, thus serving as critical stepping stones for systems biology. The information contained in the annotated genome will be more meaningful when ordered into metabolic pathways, regulatory networks and signal transduction networks to understand the cellular networks of the organism. The coding regions or open reading frames in a genome can be identified by performing similarity searches of the completed genome against databases of annotated ontologies, which provide initial clues about pathways. Some resources that help in this initial annotation are Gene Ontology<sup>1</sup> and InterPro.<sup>2</sup>

<sup>1</sup> <http://www.geneontology.org/>.

<sup>2</sup> <http://www.ebi.ac.uk/interpro/>.

### **5.3.2 *Impact of High-Throughput Experiments***

The various kinds of downstream “omics” research has led to the generation of several important resources such as gene essentiality data through the transposon site hybridisation (TraSH) method ([Sasseti et al. 2003](#)), whole-genome expression profiles facilitated by microarrays and lists of expressed proteins under different conditions through proteomics experiments, transcriptional profiles through microarray data and measured metabolite concentrations producing metabolomics data. These studies can also be carried out as comparative studies in various chemical environments. For example, the transcriptional response of each gene to a variety of conditions like genetic perturbation and response to a chemical or drug can be studied in a single experiment. It is generally believed that genes related in function (or part of the same pathway) are co-regulated and therefore exhibit similar expression profiles. The “omics” scale experimental data are an enormously useful resource for building systems level models, since they provide comprehensive parts lists of various kinds. In addition, the expression profiling data implicitly capture interactions, dependencies and influences among the various components, which are manifested in the form of correlated expression patterns. There are also a large number of important molecular biology and biochemical studies that have produced data about individual protein molecules, such as protein–protein interactions, gene knock-outs and site-directed loss-of-function and gain-of-function mutants, to name a few data types. Information from these are extremely useful in enriching the bioinformatics-based gene annotations, since they enable incorporation of more direct functional information.

## **5.4 Network Modeling and Simulation**

### **5.4.1 *Reconstruction of Mtb Metabolism***

Metabolic reconstruction is a process through which the various components of the metabolic network of a biological system, viz. the genes, proteins, reactions and metabolites that participate in metabolic activity, are identified, categorised and inter-connected to form a network. Most often, the system is a single cell of interest, and building on the genomic sequence as a scaffold, reconstructions can incorporate hundreds of reactions that approximate the entire metabolic activity of a cell. A comprehensive review of metabolic reconstruction has been published in [Reed et al. \(2006a\)](#). Metabolic reconstructions fundamentally rely on the availability of genome sequence and annotations. The reconstructed metabolic networks may be quite incomplete, if there are a lot of gaps in the annotation of the genome. It is not uncommon to find many “dead end” reactions in reconstructed metabolic networks – reactions which produce a metabolite that participates in no further downstream reactions, or reactions that consume a metabolite whose precursors are not present

in the network. With an increase in metabolomics data and improved functional annotation of genomes, these knowledge gaps are likely to become smaller. The analysis of reconstructed metabolic networks can also identify metabolic gaps and predict missing reactions required to reconcile disagreements between reconstructed metabolic networks and experimental data (Reed et al. 2006b).

For *Mtb*, the mycolic acid pathway (MAP) has been reconstructed (Raman et al. 2005) and simulated using *flux balance analysis* (FBA), a constraint-based approach for analysing metabolic networks. A mathematical abstraction of the biosynthesis of mycolic acids and the study of the pathway through FBA led to the identification of key points in the pathway and the delineation of potential drug targets. In 2007, two genome-scale reconstructions of *Mtb* were reported (Beste et al. 2007; Jamshidi and Palsson 2007), with applications in drug target identification, through the analyses of essential genes and hard-coupled reaction sets. In the following section, we discuss the basics of FBA, followed by the analysis of the MAP using FBA, as an example for metabolic reconstruction and simulation, providing insights into the biology of the pathway. The two genome-scale reconstructions are also briefly discussed.

#### 5.4.1.1 Flux Balance Analysis

One specific example of metabolic modeling using a constraint-based approach is FBA (Bonarius et al. 1997; Edwards et al. 2002; Kauffman et al. 2003; Raman and Chandra 2009), which uses linear optimisation to determine the steady-state reaction flux distribution in a metabolic network by maximising an objective function, such as ATP production or growth rate (Kauffman et al. 2003). FBA has been shown to be a very useful technique for analysis of metabolic capabilities of cellular systems (Edwards and Palsson 2000; Förster et al. 2003; Beste et al. 2007; Jamshidi and Palsson 2007). FBA involves carrying out a steady-state analysis, using the stoichiometric matrix for the system in question. An important assumption is that the cell performs optimally with respect to a metabolic function, such as maximisation of biomass production or minimisation of nutrient utilisation, on the premise that selection pressures during evolution guide systems towards optimality. Once an objective function is fixed, the system of equations can be solved to obtain a steady-state flux distribution. This flux distribution is then used to interpret the metabolic capabilities of the system.

The stoichiometric information on a metabolic system is encoded in a stoichiometric matrix, where every metabolite is represented by a row and every reaction by a column. The entries in each column correspond to the stoichiometric coefficients of the metabolites (negative for reactants and positive for products) for each reaction. The stoichiometric matrix  $S_{m \times n}$  of  $m$  metabolites and  $n$  reactions is a sparse matrix; generally, the entries are integers. The  $i$ th row defines the participation or connectivity of a particular metabolite across all metabolic reactions, and the  $j$ th column provides the stoichiometry of all metabolites in that reaction. The dynamic mass balance of the metabolic system is described using the stoichiometric matrix,

relating the flux rates of enzymatic reactions,  $\mathbf{v}_{n \times 1}$  to time derivatives of metabolite concentrations,  $\mathbf{x}_{m \times 1}$  as

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \mathbf{v} \quad (5.1)$$

$$\mathbf{v} = [v_1 \ v_2 \ \dots \ v_{n_i} \ b_1 \ b_2 \ \dots \ b_{n_{\text{ext}}}]^T \quad (5.2)$$

where  $v_i$  signifies the internal fluxes,  $b_i$  represents the exchange fluxes in the system,  $n_i$  is the number of internal metabolites and  $n_{\text{ext}}$  is the number of external metabolites in the system. At steady state,

$$\frac{d\mathbf{x}}{dt} = \mathbf{S} \mathbf{v} = 0 \quad (5.3)$$

Therefore, the required flux distribution belongs to the null space of  $\mathbf{S}$ . Since  $m < n$ , the system is under-determined and can be solved for  $\mathbf{v}$  fixing an optimisation criterion, following which the system translates into a linear programming problem:

$$\min_{\mathbf{v}} \mathbf{c}^T \mathbf{v} \quad \text{s. t.} \quad \mathbf{S} \cdot \mathbf{v} = 0 \quad (5.4)$$

where  $\mathbf{c}$  represents the objective function composition, in terms of the fluxes. Furthermore, the lower and upper bounds of the fluxes can be constrained as follows:

$$\begin{aligned} 0 &\leq v_i < \infty \\ -\infty &< b_i < \infty \end{aligned} \quad (5.5)$$

which necessitates all internal irreversible reactions to have a flux in the positive direction and allows exchange fluxes to be in either direction. Practically, a finite upper bound can be imposed, so that the problem does not become unbounded. This upper bound may also be decided based on the knowledge of cellular physiology.

FBA also has the capabilities to address the effects of gene deletions and other types of perturbations on the system. Gene deletion studies can be performed by constraining the reaction flux(es) corresponding to the gene(s) (and therefore, of their corresponding proteins(s)), to zero. Effects of inhibitors of particular proteins can also be studied in a similar way by constraining the upper bounds of their fluxes to any defined fraction of the normal flux, corresponding to the extents of inhibition.

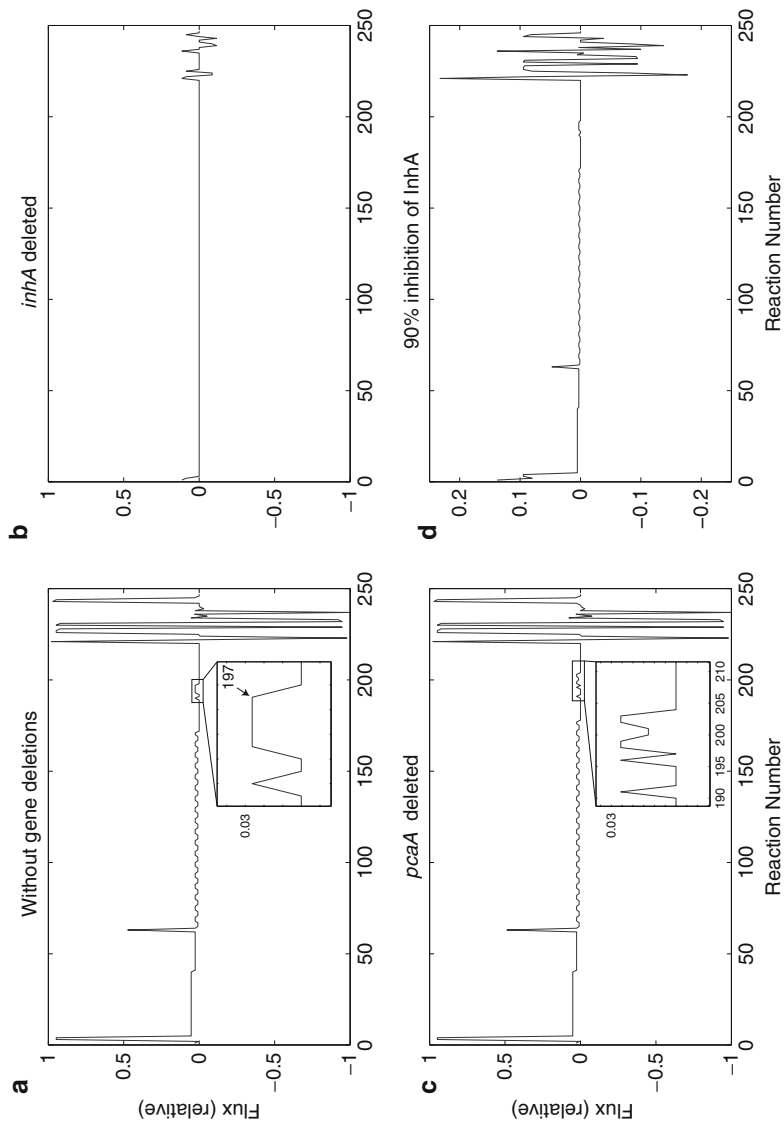
#### 5.4.1.2 Mycolic Acid Pathway

The mycobacterial cell wall is distinctive and is associated with the pathogenicity of *Mtb* (Smith 2003; Barry III et al. 1998; Dubnau et al. 2000; Glickman et al. 2000). The synthesis of mycolic acids, which are long-chain  $\alpha$ -alkyl- $\beta$ -hydroxy fatty acids, the major constituents of this protective layer, has been shown to be critical for the survival of *Mtb* (Draper and Daffé 2005).

A comprehensive stoichiometric model of MAP was built (Raman et al. 2005) using publicly available databases such as BioCyc and extensive curation of biochemical and genetic data available in literature. The model of the MAP contained 219 reactions and 197 metabolites, mediated through 28 proteins. FBA was performed on the MAP model, which provided insights into the metabolic capabilities of the pathway. For FBA, the objective function for optimisation was based on the production of various mycolates in the system, based on their relative ratios in the mycobacterial cell wall. On solving the optimisation problem, a flux distribution was obtained (Fig. 5.2).

The strength of many systems-level analyses stems from their abilities to analyse perturbations to a system; FBA can be readily applied to perform *in silico* gene deletions, whereby the effect of deleting one or more genes on the flux distribution in the system can be predicted. A systematic gene deletion study of the MAP was also carried out, as well as the inhibition of InhA by isoniazid (Raman et al. 2005). These studies provide clues about proteins essential for the pathway and hence lead to a rational identification of possible drug targets. Each of the 28 genes and hence its gene product were systematically deleted from the MAP model, one at a time, and their effect on the flux distribution was analysed. Figure 5.2b, c are examples of flux distributions upon gene deletion and correspond to deletion of *inhA* and *pcaA*, respectively. Upon deletion of *inhA*, which catalyses 21 reactions in the Fatty Acid Synthase-II system, the fluxes of almost all reactions were seen to be zero. On the other hand, upon deletion of *pcaA*, which is involved only in the production of  $\alpha$ -mycolate, the flux pattern remained largely unaltered, except for the increase in the flux corresponding to *cis*-methoxy-mycolate production. A flat flux distribution profile (of near zero) was observed upon deletion of 16 of the genes (and hence their gene-products) in the MAP model. Some other genes (as in the case of *pcaA*), when deleted, did not significantly alter the overall flux distribution, since *cis*-methoxy mycolate is produced in increased quantities, to compensate for the absence of  $\alpha$ -mycolate. Analysis of the effects of deletion of individual genes on the flux profiles of the five mycolates provided a handle to define essential and non-essential genes. Those deletions that resulted in zero or near-zero fluxes of all the mycolates were considered as essential, and the rest were considered as non-essential. A good correlation was observed for 19 genes, no experimental data were available for four genes and disagreement was seen only for five genes. High correlation with experimentally observed data about the essentiality of individual genes indicates the usefulness of the MAP model and its study using FBA.

Those genes that were classified as essential in the above analysis automatically form a first list of putative targets for anti-tubercular drugs, since their total inactivation results in loss of production of mycolic acids and hence the viability or the pathogenicity of the bacillus. However, it was reasoned that an ideal target should be essential not only in terms of the reaction it can catalyse, but also as the only protein coded by the genome that can perform the same task. Moreover, an ideal target should also have no recognisable homologue in the host system, which can in principle compete with the same drug, leading to unintended/adverse effects in the host system. Sequence analysis with the *Mtb* H37Rv and human proteomes was



**Fig. 5.2** Flux distributions obtained from flux balance analysis (FBA) using the mycolic acid pathway (MAP) model. (a) in an unperturbed state, (b) upon deletion of *inhA*, (c) upon deletion of *pcaA* and (d) upon inhibition of *InhA*. Insets in (a) and (c) refer to enlarged versions of the indicated portions. Note that the scale for (d) is different

therefore carried out for each of the proteins identified as essential from the gene deletion studies. These studies indicated that, apart from the known *InhA*, potential targets for anti-tubercular drug design are *AccD3*, *Fas*, *FabH*, *FabD*, *DesA1* and *DesA2*.

#### 5.4.1.3 Genome-Scale Metabolic Models

Two genome-scale metabolic models have been reported for *Mtb*, viz. *GSMN-TB* (Beste et al. 2007) and *iNJ661* (Jamshidi and Palsson 2007). These models are based on a careful reconstruction of *Mtb* metabolism, based on publicly available pathway databases and a survey of literature. The models consist of a stoichiometric matrix representing the metabolism of *Mtb*, the connections between the genes, proteins and reactions, as well as the various constraints on the cellular fluxes. For FBA, an objective function is used, based on the knowledge of mycobacterial biomass composition. Genome-scale metabolic models are particularly useful to understand the metabolic capabilities of organisms and for the study of gene deletions in silico, to identify critical points in the metabolism, which may be potential drug targets.

*GSMN-TB*. The genome-scale metabolic network (GSMN) of *Mtb* comprises 849 unique reactions involving 739 metabolites and 726 genes (Beste et al. 2007). The constraint-based metabolic model was calibrated by growing *M. bovis* BCG in continuous culture and the measurement of steady state growth parameters. FBA was used to calculate substrate consumption rates, which were in good agreement with experimental measurements. The objective function for FBA was the maximisation of biomass production; the biomass composition was based on the components necessary for growth in vitro (Beste et al. 2007). FBA-based in silico gene deletions were also reported, with a prediction accuracy of 78%. The model predicts that about 34% of the genes in the model are essential for growth in the minimal Middlebrook 7H10 media. FBA of the model also correctly predicts the essentiality for growth of known drug targets such as *inhA*, *embAB*, *ddlA* and *alr*. The model was able to correctly predict increased isocitrate lyase activity in slow-growing cells. The model demonstrates the predictive power of FBA-based metabolic models, which can be used to generate a number of hypotheses that may be verified experimentally. Thus, such metabolic models of pathogenic organisms provide valuable insights into the biology of the organism, paving the way for new strategies to counter disease.

*iNJ661*. Another genome-scale metabolic model for *Mtb* has been reported by Palsson and co-workers (Jamshidi and Palsson 2007). The model was used to analyse the growth of the bacterium on various in silico media. Growth rates consistent with experimental data were observed in varying media conditions. The agreement of gene essentiality predictions with experimental data was about 55%; this is due to the variability of gene expression under different conditions and the incompleteness of biological knowledge. Furthermore, hard-coupled reaction sets, which are groups of reactions that are forced to operate in unison due to the constraints in the network (arising due to mass balance and connectivity), were identified, which have application in the identification of drug targets.

Although the curation of such metabolic models is an extremely tedious process, the models are quite versatile and find use in understanding the metabolism of pathogens, and consequently in drug target identification. The importance of considering metabolism in drug design has been emphasised earlier (Cornish-Bowden and Cárdenas 2003), particularly since a key activity of drugs is to alter metabolism; many known drug targets are enzymes or receptors. Such genome-scale reconstruction studies also have their limitations. It is often difficult to accurately determine the biomass composition, which is fundamental to FBA and predictions of gene essentiality. This is particularly true in case of pathogenic organisms such as *Mtb*, where it is particularly difficult to perform experiments and determine the biomass constituents/various growth parameters; practically, it is possible to estimate such parameters from experiments with avirulent strains, such as *M. bovis*, which may not truly reflect the behaviour of *Mtb*. Furthermore, as discussed earlier, the quality of the metabolic reconstructions is limited by the availability of genome annotations. The predictions of gene essentiality also suffer from this incompleteness, as also from the incomplete definition of the biomass function, presence of unknown isozymes for a given reaction (another dimension of network incompleteness) and a failure to consider the build-up of toxic intermediates. A detailed discussion of the factors underlying incorrect in silico predictions of essential metabolic genes has been presented elsewhere (Becker and Palsson 2008). Notwithstanding these limitations, models and the simulation methodologies currently adopted still capture the metabolic structure in the cells fairly accurately and will no doubt serve as a framework to integrate newer information, as it becomes available and thus further refine the models and pose a variety of questions that may be addressed with increased confidence.

### 5.4.2 Transcriptional Analysis

The advances in microarray technology has enabled the genome-scale analysis of mRNA expression profiles in various organisms, including *Mtb*. A detailed review of genome-scale expression analyses of *Mtb* has been reported elsewhere (Waddell and Butcher 2007). A comprehensive study of the differential transcriptional response of *Mtb* to drugs and growth-inhibitory conditions has been reported earlier (Boshoff et al. 2004). A total of 430 microarray profiles were generated, which were then clustered. Agents of known mechanism of action were clustered together, while the mechanism of action of unknown agents could also be predicted (Boshoff et al. 2004). The fine clustering of genes provides insights into the metabolic response of *Mtb* to drug-induced stress, presenting a rational basis for the selection of critical metabolic targets for new anti-mycobacterials.

In another study, the response of *Mtb* to minimal inhibitory concentrations of six anti-microbials was determined, using microarray analysis to elucidate mechanisms of innate resistance in *Mtb* (Waddell et al. 2004). A common response to drug exposure which overlapped with a number of other mycobacterial stress responses,

as well as compound-specific responses were distinguished, including a number of putative transcriptional regulators and translocation-related genes. These genes may be implicated in the intrinsic resistance of *Mtb* to drugs.

It has also become possible to perform large-scale studies of gene essentiality (Sasseti et al. 2003), using microarrays. The method known as transposon site hybridisation mutagenesis uses microarrays to map sites of transposon insertions. DNA from a transposon library is isolated, and labelled probes are synthesised from promoters within the transposon. Immediately after mutagenesis, each mutant contains a single transposon insertion, and the library contains mutations in each gene in the genome. After a growth phase, mutants harbouring insertions in genes that are required for survival are lost from the library. A TraSH “insertion probe” is generated from the selected library, comprising only those sequences complementary to genes that contain insertions in the selected library. A genomic probe comprising randomly labelled chromosomal DNA will hybridise to every gene represented on the array. Spots that hybridise to the genomic probe, but not to the insertion probe, represent genes that are required for mycobacterial growth (Sasseti et al. 2003).

The analysis of the bacterial transcriptional response to infection can elucidate the physiological state of the infecting bacteria, bacterial mechanisms to counter infection, as well as the micro-environments encountered by the bacteria during the course of infection (Waddell et al. 2007). Whole-genome transcriptional profiling of both host and pathogen in ex vivo, animal model and human disease contexts have been reviewed in Waddell et al. (2007). More recently, an RNA amplification strategy that has potential to throw light on host–pathogen interactions has also been reported (Waddell et al. 2008). The expression of *Mtb* genes in macrophages has also been studied (Schnappinger et al. 2003) by analysing RNA isolated from infected murine macrophages using microarrays. In the macrophages, 454 induced genes and 147 repressed *Mtb* genes were identified (compared to broth cultures), termed as the “differential intraphagosomal transcriptome”. Integrating such genome-scale transcriptional analyses, which provide a wealth of data, can aid in improving the understanding of TB disease progression.

#### 5.4.2.1 Transcriptional Regulatory Networks in *Mtb*

Balázsi and co-workers have reported a large transcriptional regulatory networks (TRN) in *Mtb* characterising the temporal response of this network during adaptation to stationary phase and hypoxia, using published microarray data (Balázsi et al. 2008). The TRN principally consists of gene regulatory interactions from literature as well MtbRegList (Jacques et al. 2005). The network was further expanded based on orthology with *Escherichia coli*. All *Mtb* operons was also incorporated, based on the assumption that transcription factor (TF) binding to the promoter region affects the expression of all genes within an operon. The TRN comprises 783 nodes corresponding to *Mtb* genes and their protein products, with 937 links corresponding to 45 TFs directly regulating the expression of target genes. Significantly, 29 of

these 45 TFs (auto-)regulate their own expression. Gene pairs such as Rv2358-furB, Rv1404-Rv1931c and mprA-sigE participate in two-gene feedback loops. A distinct set of transcriptional sub-networks affected early and late during adaptation to hypoxia and stationary phase were identified, illustrating a progressive shift of modular network response to growth arrest. Most of the sub-networks were affected in both conditions, suggesting that a general condition-independent repertoire of transcriptional modules is used in *Mtb* growth arrest (Balázsi et al. 2008).

Studies such as this hold the key to unravelling the mechanisms of mycobacterial persistence, which is a critical problem in mycobacterial infection, where the bacteria enter a non-replicating state, insensitive to anti-mycobacterial drugs.

### 5.4.3 Analysis of the *Mtb* Interactome

Protein–protein interactions are extremely important in orchestrating the events in a cell. They form the basis for several signal transduction pathways in the cell, as well as various transcriptional regulatory networks. The need to understand protein structure and function has been a critical driving force for biological research in the recent decades.

Genome-wide functional linkages between proteins can be inferred from high throughput experimentation or from computational analyses. Eisenberg and co-workers have reported genome-wide functional linkages in *Mtb* (Strong et al. 2003), inferred by computational methods based on genomic context, such as the Rosetta Stone, which is based on domain fusion (Marcotte et al. 1999), Phylogenetic profile, based on co-occurrence of proteins across genomes (Pellegrini et al. 1999), Operon and Conserved Gene Neighbour, based on the proximity of genes on the chromosome across several genomes (Dandekar et al. 1998). By clustering proteins with similar functional linkage profiles, it is possible to infer the function of uncharacterised proteins and identify functionally linked gene clusters across the proteome. This study once again demonstrates the utility of a genome-scale analyses vis-à-vis analyses of individual protein interactions/functional linkages. Such protein–protein interaction maps also find utility in drug target identification (Verkhedkar et al. 2007; Raman et al. 2008) and in the analysis of resistance pathways (Raman and Chandra 2008), as will be discussed in later sections. Various concepts from graph theory have been applied to study biological networks (Barabási and Oltvai 2004). Many of the highly connected proteins in protein interaction networks, referred to as “hubs”, have been shown to be critical for cellular function (Jeong et al. 2001); such hub proteins also represent potential drug targets (Verkhedkar et al. 2007).

Although such protein–protein functional linkages enable a wide variety of analyses, in some ways they represent an over-simplified static view of the dynamic interactions in the cell. Presently, it is difficult to estimate the parameters governing these interactions, such as association and dissociation constants, which govern many aspects of cellular function, particularly signalling. High-throughput studies for identifying protein–protein interactions, such as the yeast

two-hybrid assay (Fields and Song 1989), despite their numerous advantages and versatility, produce a number of false positives and false negatives, although these are being addressed a number of recent advances. A comprehensive overview of the systems biology applications and limitations of the yeast two-hybrid assay is presented elsewhere (Brückner et al. 2009). Computational methods to predict functional linkages also suffer from false positives and negatives, but these can be addressed by considering consensus predictions from multiple methods. The STRING database (Von Mering et al. 2007) considers predictions based on multiple methods as well as experimental data and assigns a confidence score to each interaction. With a refinement of both computational and experimental techniques to delineate protein–protein interactions, the quality of constructed interactomes is likely to improve significantly in the future, enabling analyses with greater confidence.

## 5.5 Target Identification

Drug discovery has itself witnessed a paradigm shift from the traditional medicinal chemistry-based ligand-oriented drug discovery approaches to rational drug target identification and target-driven lead discovery, by targeting the molecular mechanisms of disease. Traditionally, targets have been identified through knowledge of the function of individual protein molecules, where their function has been well characterised. Potential targets thus identified are generally taken through a validation process involving whole-cell or animal experiments, gene knock-outs or site-directed mutagenesis that lead to loss-of-function phenotypes. Target validation is one of the critical steps in drug discovery, where a lot of time and money is spent in the pharmaceutical industry. The need for systematic and large-scale validation in the post-genomic era has led to the use of computational methods for validation (Raman et al. 2007).

A number of studies have been carried out using various experimental methods to identify drug targets in *Mtb* (Mdluli and Spigelman 2006). Attempts have also been made for the same purpose, based on sequence comparisons of metabolic enzymes (Anishetty et al. 2005), and using various features such as metabolic choke-points at the systems-level (Hasan et al. 2006). The wealth of information available from the genome sequence, as well as metabolic and protein interaction networks can be analysed to identify potential drug targets in *Mtb*. In this section, we discuss how systems biology concepts and understanding the microbe as a whole open up new opportunities for computational target identification.

### 5.5.1 Multi-Level Target Identification Pipeline: TargetTB

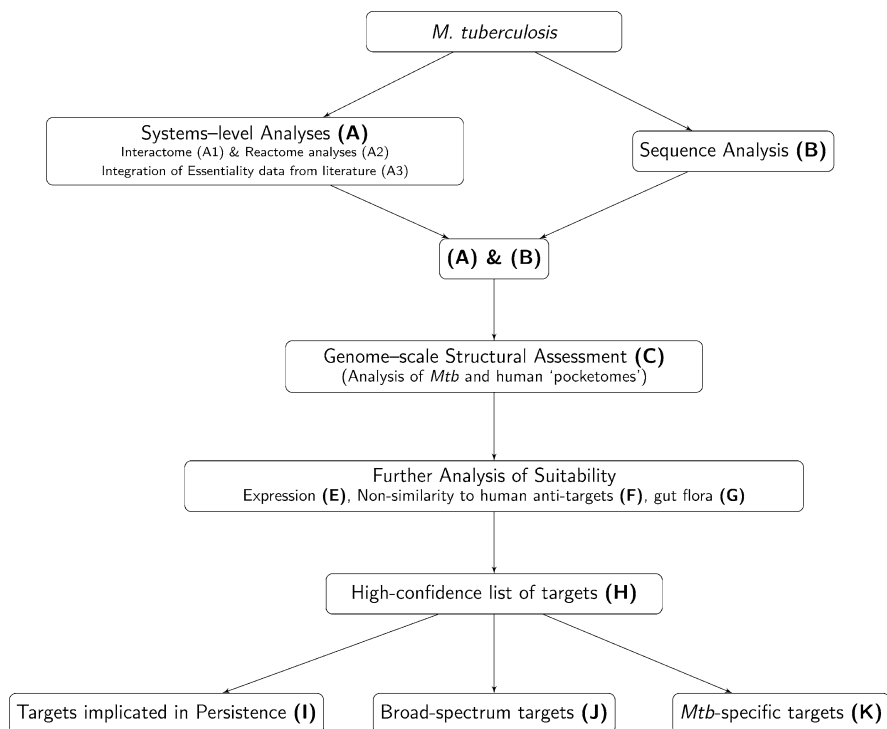
It is now well established that better insights into biological systems may be obtained by considering large-scale systems-level models, since biological systems

are complex networks of many processes. The conventional method of focussing on a single protein at a time, however important the protein may be, would mean losing perspective of its larger context and hence may not provide the right answers, especially in drug discovery. Broader insights about the appropriateness of a potential target can be obtained by considering pathways and whole-system models relevant to that disease. For example, an enzyme that may be identified as a good target for a particular disease may not actually be critical or essential, when viewed in the context of the entire metabolism in the cell. Analysing systems-level models can help in assessing criticality of the individual proteins by studying any alternate pathways and mechanisms that may naturally exist to compensate for the absence of that protein.

An integrated analysis of *Mtb* at various levels – metabolic reactions, protein–protein interactions, protein sequences and structure – can provide a more rational handle to identify drug targets. Illustrating this, a comprehensive in silico target identification pipeline for *Mtb*, targetTB (Raman et al. 2008), has been reported, which can also be used as a general framework for in silico target identification. The analyses are focused at a systems-level, based on network analyses and FBA. The pipeline incorporates (a) a network analysis of the protein–protein interactome, (b) an FBA of the reactome, (c) experimentally derived phenotype essentiality data, (d) sequence analyses and (e) a structural assessment of targetability using novel algorithms. Using FBA and network analysis, proteins critical for survival of *Mtb* are first identified, followed by comparative genomics with the host, finally incorporating a novel structural analysis of the binding sites to assess the feasibility of a protein as a target. Further pruning of the chosen targets was done based on (f) analysis of expression of suggested target proteins, based on available expression data and (g) non-similarity to gut flora proteins as well as (h) “anti-targets” in the host, leading to the identification of 451 high-confidence targets. Through phylogenetic profiling against 228 pathogen genomes, shortlisted targets have been further explored to identify broad-spectrum antibiotic targets, while also identifying those specific to TB. Targets that address (i) mycobacterial persistence and (j) drug resistance mechanisms are also analysed.

Besides essentiality to the pathogen, an ideal target should have several other properties such as non-similarity with human proteins whose inhibition could lead to potential adverse drug effects, an aspect that has been analysed at multiple levels in this study (see Fig. 5.3).

The simplest level of course is to check for sequence similarity of the target being queried with all the proteins in the human proteome. However, sequence filtering while important cannot be the sole criteria for identifying high quality targets, since two proteins that are considerably dissimilar in their sequences could have very similar binding sites (Ramachandraiah and Chandra 2000; Vinod et al. 2006). Thus, while sequence similarity very often leads to structural and hence functional similarity, it is not a necessary condition for two proteins to have similar ligand-binding profiles.



**Fig. 5.3** The targetTB target identification pipeline. The flowchart depicts the order in which the entire proteome of *Mtb* is considered and analysed at different layers. “A” refers to the systems level studies, which includes A1, for network analysis of the interactome; A2, for FBA of the reactome; and A3, for genome-scale essentiality data determined experimentally as reported by Sassetti et al. (2003). Those proteins that passed these filters are indicated as “A”, and combined with the results of sequence analysis (A), to derive those that passed both filters (depicted as “A&B”). These were then taken through Filter C, referring to the structural assessment filter, yielding the list of 622 proteins as the D-List (A&B&C). Further steps of filtering are indicated in the smaller funnel as E (expression under various conditions), F (non-similarity to anti-targets) and G (non-similarity to gut flora proteins). Those proteins that pass all the six levels of filtering (indicated as D&E&F&G) form the H-List comprising 451 targets. Additional filters I, J and K used for analysing the H-List are also indicated

**Genome-scale structural assessment.** In the process of target identification, the critical aspect of a good target is to have a binding site in the target protein that is sufficiently different from that of any host protein. It is important to consider specificity at the binding site, hence the molecular recognition level, since a given drug should be available in intended quantities to the desired target. At the same time, a given drug should ideally not exhibit unintended recognition by some other host protein, so that adverse effects due to unanticipated functional manipulation of other host proteins can be avoided. For this purpose, it is important to study the possible binding profile of a given drug to all those proteins to which it is likely to be exposed. Towards this goal, possible pockets in the set of *Mtb* and human structures

were first identified, using PocketDepth, a validated algorithm that was recently developed (Kalidas and Chandra 2008). All such putative pockets were tested for certain criteria such as size and volume, retaining only those that were likely to bind to small molecules. The filtered pockets from preliminarily short-listed targets from *Mtb* were then screened for similarity against pockets from the human proteins, which involved over 245 million comparisons, using PocketMatch, another recently developed site-matching algorithm (Yeturu and Chandra 2008). From this, 145 putative targets were eliminated due to high similarity with one or more human proteins. Some examples of molecules that have failed at this stage are DdlA, GyrB, AftA and AlrA. It must be noted that some of these were ranked as high priority targets by other studies that did not consider the structural aspect explicitly, again emphasising the need for structural level analysis. Eliminating those proteins with high similarity to proteins in the gut flora also helps in ultimately reducing the risk of side effects.

The last stages of filtering and post-identification analyses resulted in identifying two categories of targets: broad-spectrum targets and *Mtb*-specific targets. It is necessary to identify targets in both the categories, since they are required in different situations. *Mtb*-specific targets are believed to be safer since they would not lead to many organisms developing resistance against the drugs of such targets. Broad-spectrum targets, on the other hand, would be extremely useful when multiple infections co-exist or in some cases where a specific diagnosis is not possible. A comprehensive phylogenetic analysis of the short-listed targets against 228 different pathogenic genomes has been carried out, to identify broad-spectrum targets. Identification of pathways and proteins involved in generating drug resistance and then targeting them simultaneously as co-targets along with the primary broad-spectrum targets would reduce the risk of drug resistance significantly, making many more molecules accessible for therapeutic intervention.

### 5.5.1.1 Importance of Systems-Based Approaches

The pipeline described shows how systems biology methods can be used to obtain significant insights into essentiality, identifying possible lists of essential proteins and of course understanding reasons for their essentiality as well. The study described here demonstrates the usefulness of such insights in target identification for tuberculosis and how they can be integrated along with other canonical lines of investigation such as sequence and structural analyses of the individual molecules. The pipeline developed provides rational schema for drug target identification that are likely to have high rates of success, which is expected to save enormous amounts of money, resources and time in the drug discovery process. A thorough comparison with previously suggested targets in the literature demonstrates the usefulness of the integrated approach used in the study, highlighting the importance of systems-level analyses in particular (Raman et al. 2008). The method has the potential to be used as a general strategy for target identification and validation and hence significantly impacts most drug discovery programmes.

### 5.5.2 *Disruption of Metabolism*

It has been said that drug design has often not included the idea that what cells do is metabolism, and a major thing drugs are supposed to do is to alter metabolism: of the 500 well-known targets, 30% are enzymes and 45% are receptors (Cornish-Bowden and Cárdenas 2003). Therefore, it is quite important to consider metabolism during drug design and drug target identification. Given this, disrupting mycobacterial metabolism to the point of destruction would be a more useful approach than to consider one metabolic target at a time. A recent study highlights the use of a protein–protein influence network derived from metabolic linkages to identify combinations of proteins, which when simultaneously inhibited can together disrupt bacterial metabolism to a significant extent, thus ensuring bacterial clearance (Raman et al. 2009). An FBA of these identified combinations indicate that metabolism has indeed been disrupted. With key proteins in the network, multiply targeted, the chances of recovery by the bacilli and emergence of drug resistance would also be hampered in a major way. Targeting multiple points in a metabolic pathway can be a useful strategy in drug design in general, perhaps explaining why combination therapy is popular. A report in literature highlights the possibility of “natural” crude drugs acting on multiple targets with multiple mechanisms, attributing their success to this plurality (Csermely et al. 2005).

### 5.5.3 *Tackling Resistance in *Mtb**

A major problem with the current chemotherapeutic agents for TB is the emergence of drug resistance. Although several approaches have been explored to counter resistance, there has been limited success due to a lack of understanding of how resistance emerges in bacteria upon drug treatment.

A proteome-scale network of protein–protein associations in *Mtb* has been used, to discover possible pathways that may be responsible for generating drug resistance (Raman and Chandra 2008). The protein–protein interactome of *Mtb* enables a novel formulation of the problem of drug resistance and forms a first step towards countering drug resistance at the drug discovery stage itself. In particular, the questions such as: (a) how does the information flow from the drug target to the resistance machinery, and (b) how do targets differ in their propensities for triggering resistance, can be addressed.

A genome scale protein–protein interaction network for *Mtb* H37Rv was derived from the STRING database (Von Mering et al. 2007). A set of proteins involved in both intrinsic and extrinsic drug resistance mechanisms were identified from literature. Shortest paths from different drug targets to the set of resistance proteins in the protein–protein interactome were computed, to derive a sub-network relevant to study emergence of drug resistance. The shortest paths were then scored and ranked based on (a) drug-induced gene upregulation data, from microarray experiments reported in literature, for the individual nodes and (b) edge-hubness, a

network parameter that signifies centrality of a given edge in the network. High-scoring paths, which contain “central” proteins up-regulated on exposure drugs, indicate most plausible pathways for the emergence of drug resistance. Different targets appear to have different propensities for four drug resistance mechanisms, giving rise to a very important direction to explore in drug discovery.

The study leads to the identification of possible pathways for drug resistance, providing novel insights into the problem of resistance. A new concept of “co-targets”, to counter resistance by simultaneously inhibiting a protein responsible for resistance, along with the intended target of the drug, has been proposed to counter mycobacterial drug resistance. RecA, Rv0823c, Rv0892 and DnaE1 were among the best examples of co-targets for combating drug resistance in TB (Raman and Chandra 2008). This approach is also inherently generic, likely to significantly impact drug discovery.

## 5.6 Interface with the Host: Modeling Host–Pathogen Interactions

The establishment of any infection is contingent upon the interplay of virulence mechanisms of a pathogenic organism, the defence mechanisms of the host as well as the counter-defence of either organism. A comprehensive understanding of the mechanisms of host–pathogen interactions can aid in the identification of the critical points for countering infection. Although a comprehensive mechanistic model of host–pathogen systems is still not available, several approaches have been undertaken to identify and model host–pathogen interactions. These range from simpler models for the prediction of protein–protein interactions between the host and pathogen, to complex models for the signal transduction networks and Boolean network models of immunological components of the interplay of various mechanisms of attack and defence in the host and pathogen.

### 5.6.1 Response Networks

Response network analysis involves the analysis of experimental data such as gene expression profiles, in the context of biological networks. Superposing network information with experimental data, networks representing the best system response according to the tested experimental conditions are identified (Forst 2006). Siegel and co-workers integrate expression data with molecular interaction data to identify active sub-networks, or, connected regions of the network showing significant changes in expression (Ideker et al. 2002). Forst and co-workers also explore the differential network expression during response of *Mtb* to stress (induced by hydrogen peroxide) and drugs such as isoniazid, using concepts from graph theory (Cabusora et al. 2005). The expression data of known stress responders and DNA repair genes in *Mtb* were used to construct a generic stress response sub-network. This was then

compared to similar networks constructed from data obtained from subjecting *Mtb* to various drugs; this analysis helps to distinguish between generic stress response and specific drug response, which can be exploited in drug discovery. With a growth in microarray data, it is possible to extend these ideas to the host–pathogen interactome networks. Genes that are selectively expressed during infection may be more likely to be involved in virulence.

### 5.6.2 Mechanistic Models of Immune System Dynamics

Kirschner and co-workers have worked on several mathematical models for the interaction of *Mtb* with the human immune system (Wigginton and Kirschner 2001; Marino and Kirschner 2004; Segovia-Juarez et al. 2004; Marino et al. 2007a). These mathematical models use differential equations encapsulating the interactions between various host cells, cytokines and the pathogen. Comprehensive reviews of mathematical models of *Mtb* infection and its interactions with the human immune system have been published elsewhere (Kirschner and Marino 2005; Young et al. 2008). A virtual model of the immune response to *Mtb* that characterises the cytokine and cellular network that is operational during TB infection has been reported (Wigginton and Kirschner 2001). The dynamics of the various model components such as macrophage and cytokines are described using differential equations. Using this model, the parameters governing the behaviour of the system towards the different outcomes have been identified. The study concludes that factors affecting macrophage functions (such as activation, infection and bactericidal capabilities) as well as effector T cell functions (cytotoxicity from CD4<sup>+</sup> T cells as well as other cells such as CD8<sup>+</sup> T cells) must achieve a balance to control infection. Virtual deletion and depletion experiments have also been performed to confirm these hypotheses. The model has been further extended to a two compartmental model capturing the important processes of cellular activation and priming that occur between the lung and the nearest draining lymph node. The model is able to reproduce typical disease progression scenarios including primary infection, latency or clearance (Marino and Kirschner 2004). Agent-based models for simulating granuloma formation have also been reported (Segovia-Juarez et al. 2004).

Marino and Kirschner have developed a model which shows that delays in either dendritic cell migration to the draining lymph node or T-cell trafficking to the site of infection can alter the outcome of *Mtb* infection (Marino et al. 2004). A mathematical model of immune response to *Mtb* in the lungs, exploring the role CD8<sup>+</sup> T cells, has also been developed (Sud et al. 2006). Ray and Kirschner have also developed a mathematical model comprising several differential equations describing macrophage biochemical processes based on three functional modules, viz. activation, killing and iron regulation (Ray and Kirschner 2006). They suppose the requirement of multiple activation signals for the macrophage to overcome the quiescent state. While the innate immune response develops first occurring on the order of minutes and hours, adaptive immunity follows later occurring on the order of days

or weeks. Each has an inherent delay in their development, and this timing may be crucial in determining success or failure in clearing the pathogen. A general model of the twofold immune response, specifically to intracellular bacterial pathogens, incorporating mathematical delays for both innate and adaptive immune response has been developed (Beretta et al. 2007).

The role of tumour necrosis factor (TNF- $\alpha$ ) in protection against the tubercle bacillus in both active and latent infection has also been modeled, providing insights into the role of TNF- $\alpha$  in TB pathology and control (Marino et al. 2007b). The model consists of non-linear differential equations describing the dynamics of macrophage, T cells, cytokines and bacteria. The effect of TNF- $\alpha$  and IFN- $\gamma$  signalling on activation of the macrophage during *Mtb* infection has also been analysed using a mathematical model (Ray et al. 2008). Each component of the model, such as TNF- $\alpha$ , IFN- $\gamma$  and nitric oxide (NO), is represented as a continuous entity in an ordinary differential equation. Using the model, it has been shown that negative feedback from production of nitric oxide, the key mediator of mycobacterial killing, which typically optimises macrophage responses to activating stimuli, may reduce effective killing of *Mtb*.

### 5.6.3 Boolean Modeling of *Mtb*-Human Interactions

The roots of Boolean network modeling may be traced to as early as 1969, when Kaufmann described the use of such models for studying cellular control processes (Kauffman 1969). Another insightful exposition of Boolean network theory for modeling genetic circuits was given later by Thomas (1973). Boolean network models have been used successfully to predict the expression pattern of the segment polarity genes in *Drosophila melanogaster* (Albert and Othmer 2003). Brahmachari and co-workers have applied Boolean network modeling to analyse a neurotransmitter pathway implicated in schizophrenia (Gupta et al. 2007). Albert and co-workers have applied Boolean networks for the modeling of host–pathogen interactions in *Bordetella* (Thakar et al. 2007).

Boolean network models are composed of various nodes, representing important components or processes in the system. The state of each node in the network can be either “on” (*true*) or “off” (*false*), a qualitative description of the concentration or activity. Boolean network representations involve transfer functions that encode the interactions between the various states. Transfer functions define a discrete dynamic system, using logical constructs such as “AND”, “OR” and “NOT”. For example, activations can be represented by an “OR” operator, while an inhibition can be encoded for, by an “AND NOT” operator. When more than one of the components need to be present concurrently, to cause an activation, an “AND” can be used. Each iteration of simulation determines the evolution of the state of nodes.

One of our latest studies involves building a multi-level model of host–pathogen interactions in TB, based on an extensive survey of various experiments reported in literature, accounting for the innate and adaptive immune responses of the host, as

well as the various defence mechanisms of the pathogen (Raman 2008; Raman et al. 2010). The complex regulation by the various cytokines present in the cell has also been encoded in the model. The model contains 75 nodes, about one-fourth of them relating to bacterial components, the rest being components of the human immune system. Boolean transfer functions describe the relationships between the nodes. For example, the state of activated dendritic cells in the system could be described as follows:

$$\text{Activated\_Dendritic\_cells}^* = (\text{Dendritic\_cells and Bacteria}) \text{ or } \text{Activated\_phagocytic\_cells} \text{ or } (\text{Dendritic\_cells and Bacteria and (Th1RC or Th2RC)}).$$

This is based on the knowledge that immature DCs, upon stimulation by bacteria, get activated and mature in the lymph nodes; the activation of dendritic cells may also be aided by activated phagocytic cells and cytokines produced in T helper cells (Th1 or Th2).

Virtual deletion experiments have been performed, where one or more components of the system are removed and the response of the system to this perturbation is analysed. Disabling processes such as phagocytosis and phagolysosome fusion or important cytokines such as TNF- $\alpha$  and IFN- $\gamma$  greatly impaired bacterial clearance, while removing cytokines such as IL-10 alongside bacterial defence proteins such as SapM greatly favoured bacterial clearance. The propensity of the tubercle bacillus to persist is highlighted in the simulations. Studies of this nature are useful to identify key points in the human immune response as well as the components critical for the elimination of bacteria. An overall understanding of the interplay of the various mechanisms in host–pathogen interaction lays an excellent foundation for tackling the disease.

## 5.7 Future Perspectives

Systems biology signals a departure from the now common view in drug discovery of “single target, one drug, lone therapeutic indication”. Targeting a broader range of related biological structures should result in compounds that have common structural and functional properties, and common mechanisms of action, ultimately creating the potential for the application of a therapeutic to multiple diseases by targeting common pathways implicated in pathogenesis (Davidov et al. 2003). The culmination of systems modeling lies in the modeling of complete systems, accounting for all component reactions, the localisation of these components and their interactions. The interaction between these organelles or compartments and the interface with the physical world, in terms of external temperature, pH and other effects becomes more relevant in the highest levels of biological organisation. Computational models of human physiology come into play both to relate to whole animal models used in traditional pharmacology and more importantly to build integrated data-driven models that can be refined to mimic the human physiology more closely. The Physiome project (Hunter and Borg 2003) (<http://www.physiome.org/>) is one

such effort aimed at describing the human organism quantitatively, to understand key elements of physiology and pathophysiology. The salient features of the project are the databasing of physiological, pharmacological and pathological information on humans and other organisms and integration through computational modeling.

Molecular level understanding of the processes involved in the pharmacokinetics, bio-availability and toxicity is still very poor. With the current rate of advances in systems biology, we can also expect significant enhancements in pathway models, process models and indeed in entire system models, both in terms of mathematically representing such complex phenomena as well as in terms of mimicking and simulating the biological events.

While several advances have been made in modeling the host–pathogen interplay, there still remains a lot to be explored. Accurate mechanistic models of host–pathogen systems can give reliable insights into complicated phenomena. However, such models are often limited in their scope. On the other hand, systems-level models give a much better holistic view of the interplay, at the expense of some accuracy. Large-scale systems-level models of host–pathogen interactions, integrating information from various levels of abstraction, would be of immense use in understanding processes of infection and developing strategies for combating disease. The future of host–pathogen systems modeling holds promise for uncovering the molecular bases of disease and consequently aids in the discovery of novel therapies. We can also envisage that the use of pharmacogenomics and tailor-made medicines could be distinct possibilities in the near future.

Experimentation and computational modeling must be used in complement, each deriving benefits from the other. Computational modeling can be used to generate novel hypotheses, which can then be used to guide experimentation. Experimental verification or validation of a model can render it much more useful, as more reliable predictions can be made, on the strength of its proven validity. Systems biology approaches are also likely to impact development of molecular level pharmacokinetic and pharmacodynamic models for individual drugs, to provide comprehensive profiles of drug actions. Development of comprehensive systems-level models that encode most of the features of a system will enable a better understanding of drug toxicity and hence eliminate poor candidates early in the discovery pipeline. Insights that systems-level models can ultimately translate into more rational and personalised therapeutic intervention strategies in clinical practice. Thus, the stage is set for the integration and application of skills from mathematics, computer science and engineering disciplines, to address complex problems in biology and drug discovery, in a big way.

## References

- R. Albert and H. G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol*, 223(1):1–18, 2003
- P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972

- S. Anishetty, M. Pulimi, and G. Pennathur. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput Biol Chem*, 29(5):368–378, 2005
- G. Apic, T. Ignjatovic, S. Boyer, and R. B. Russell. Illuminating drug discovery with biological pathways. *FEBS Lett*, 579(8):1872–1877, 2005
- G. Balázsi, A. P. Heath, L. Shi, and M. L. Gennaro. The temporal response of the *Mycobacterium tuberculosis* gene regulatory network during growth arrest. *Mol Syst Biol*, 4:225, 2008
- A.-L. Barabási and Z. N. Oltvai. Network biology: Understanding the cell’s functional organization. *Nat Rev Genet*, 5(2):101–113, 2004
- C. E. Barry III, R. E. Lee, K. Mdluli, A. E. Simpson, B. G. Schroeder, R. A. Slayden, and Y. Yuan. Mycolic acids: Structure, biosynthesis and physiological functions. *Prog Lipid Res*, 37:143–179, 1998
- S. A. Becker and B. Ø. Palsson. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst Biol*, 2:14, 2008
- E. Beretta, M. Carletti, D. E. Kirschner, and S. Marino. Stability analysis of a mathematical model of the immune response with delays, In: *Mathematics for life science and medicine*, pages 177–206. Springer, Berlin, 2007
- D. J. V. Beste, T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. E. Bushell, P. Wheeler, S. Klant, A. M. Kierzek, and J. McFadden. GSMN-TB: A web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol*, 8:R89, 2007
- H. P. J. Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends Biotechnol*, 15(8):308–314, 1997
- S. Bonora and G. Di Perri. Interactions between antiretroviral agents and those used to treat tuberculosis: Clinical pharmacology of antiretroviral drugs. *Curr Opin HIV & AIDS*, 3:306–312, 2008
- H. I. Boshoff, T. G. Myers, B. R. Copp, M. R. McNeil, M. Wilson, and C. E. Barry III. The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: Novel insights into drug mechanisms of action. *J Biol Chem*, 279(38):40174–40184, 2004
- A. Brückner, C. Polge, N. Lentze, D. Auerbach, and U. Schlattner. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*, 10(6):2763–2788, 2009
- L. Cabusora, E. Sutton, A. Fulmer, and C. V. Forst. Differential network expression during drug and stress response. *Bioinformatics*, 21(12):2898–2905, 2005
- J.-C. Camus, M. J. Pryor, C. Medigue, and S. T. Cole. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, 148(10):2967–2973, 2002
- B. L. Claus and D. J. Underwood. Discovery informatics: Its evolving role in drug discovery. *Drug Discov Today*, 7:957–966, 2002
- S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogerand, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393:537–544, 1998
- A. Cornish-Bowden and M. L. Cárdenas. Metabolic analysis in drug design. *C R Biol*, 326(5): 509–515, 2003
- D. C. Crick, S. Mahapatra, and P. J. Brennan. Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*. *Glycobiology*, 11:107R–118R, 2001
- P. Csermely, V. Ágoston, and S. Pongor. The efficiency of multi-target drugs: The network approach might help drug design. *Trends Pharmacol Sci*, 26:178–182, 2005
- T. Dandekar, B. Snel, M. A. Huynen, and P. Bork. Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem Sci*, 23(9):324–328, 1998
- E. J. Davidov, J. M. Holland, E. W. Marple, and S. Naylor. Advancing drug discovery through systems biology. *Drug Discov Today*, 8(4):175–183, 2003
- J. Doyle. Computational biology. Beyond the spherical cow. *Nature*, 411(6834):151–152, 2001

- P. Draper and M. Daffé. The cell envelope of *Mycobacterium tuberculosis* with special reference to the capsule and outer permeability barrier. In: Stewart T. Cole, Kathleen D. Eisenach, David N. McMurray, and William R. Jacobs Jr., editors, *Tuberculosis and the tubercle bacillus*, pages 261–273. American Society of Microbiology Press, 2005
- E. Dubnau, J. Chan, C. Raynaud, V. P. Mohan, M. A. Lanéelle, K. Yu, A. Quémar, I. Smith, and M. Daffé. Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. *Mol Microbiol*, 36(3):630–637, 2000
- J. S. Edwards and B. Ø. Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA*, 97(10):5528–5533, 2000
- J. S. Edwards, M. W. Covert, and B. Ø. Palsson. Metabolic modelling of microbes: The flux-balance approach. *Environ Microbiol*, 4(3):133–133, 2002
- S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989
- C. V. Forst. Host-pathogen systems biology. *Drug Discov Today*, 11(5–6):220–227, 2006
- J. Förster, I. Famili, P. Fu, B. Ø. Palsson, and J. Nielsen. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, 13(2):244–253, 2003
- M. S. Glickman, J. S. Cox, and W. R. Jacobs Jr. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell*, 5(4):717–727, 2000
- S. Gupta, S. S. Bisht, R. Kukreti, S. Jain, and S. K. Brahmachari. Boolean network analysis of a neurotransmitter signaling pathway. *J Theor Biol*, 244(3):463–469, Feb 2007
- S. Hasan, S. Dugel, P. S. Rao, and M. Schreiber. Prioritizing genomic drug targets in pathogens: Application to *Mycobacterium tuberculosis*. *PLoS Comput Biol*, 2(6):e61, 2006
- P. J. Hunter and T. K. Borg. Integration from proteins to organs: The Physiome project. *Nat Rev Mol Cell Biol*, 4(3):237–243, 2003
- T. Ideker, O. Ozier, B. Schwikowski, and A. F. Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18 Suppl 1:S233–S240, 2002
- P.-E. Jacques, A. L. Gervais, M. Cantin, J.-F. Lucier, G. Dallaire, G. Drouin, L. Gaudreau, J. Goulet, and J. Brzezinski. MtbRegList, a database dedicated to the analysis of transcriptional regulation in *Mycobacterium tuberculosis*. *Bioinformatics*, 21(10):2563–2565, 2005
- N. Jamshidi and B. Ø. Palsson. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol*, 1:26, 2007
- H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001
- Y. Kalidas and N. Chandra. Pocketdepth: A new depth based algorithm for identification of ligand binding sites in proteins. *J Struct Biol*, 161(1):31–42, 2008
- K. J. Kauffman, P. Prakash, and J. S. Edwards. Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5):491–496, 2003
- S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*, 22(3):437–467, 1969
- D. Kirschner and S. Marino. *Mycobacterium tuberculosis* as viewed through a computer. *Trends Microbiol*, 13(5):206–211, 2005
- E. M. Marcotte, M. Pellegrini, H.-L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. Detecting protein function and protein-protein interactions from genome sequences. *Science*, 285(5428):751–753, 1999
- S. Marino and D. E. Kirschner. The human immune response to *Mycobacterium tuberculosis* in lung and lymph node. *J Theor Biol*, 227(4):463–486, 2004
- S. Marino, S. Pawar, C. L. Fuller, T. A. Reinhart, J. L. Flynn, and D. E. Kirschner. Dendritic cell trafficking and antigen presentation in the human immune response to *Mycobacterium tuberculosis*. *J Immunol*, 173(1):494–506, 2004
- S. Marino, E. Beretta, and D. E. Kirschner. The role of delays in innate and adaptive immunity to intracellular bacterial infection. *Math Biosci Eng*, 4(2):261–288, 2007a

- S. Marino, D. Sud, H. Plessner, L. P. Lin, J. Chan, J. L. Flynn, and D. E. Kirschner. Differences in reactivation of tuberculosis induced from anti-TNF treatments are based on bioavailability in granulomatous tissue. *PLoS Comput Biol*, 3(10):1909–1924, 2007b
- K. Mdluli and M. Spigelman. Novel targets for tuberculosis drug discovery. *Curr Opin Pharmacol*, 6(5):459–467, 2006
- P. Nunn, B. Williams, K. Floyd, C. Dye, G. Elzinga, and M. Raviglione. Tuberculosis control in the era of HIV. *Nat Rev Immunol*, 5(10):819–826, 2005
- J. A. Papin, T. Hunter, B. Ø. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6(2):99–111, 2005
- M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci USA*, 96(8):4285–4288, 1999
- G. Ramachandraiah and N. Chandra. Sequence and structural determinants of mannose recognition. *Proteins*, 39(4):358–364, 2000
- K. Raman. *Systems-level modelling and simulation of Mycobacterium tuberculosis: Insights for drug discovery*. PhD thesis, Indian Institute of Science, Bangalore, 2008
- K. Raman and N. Chandra. *Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol*, 8:234, 2008
- K. Raman and N. Chandra. Flux balance analysis of biological systems: Applications and challenges. *Brief Bioinform*, 10(4):435–449, 2009
- K. Raman, P. Rajagopalan, and N. Chandra. Flux balance analysis of mycolic acid pathway: Targets for anti-tubercular drugs. *PLoS Comput Biol*, 1(5):e46, 2005
- K. Raman, Y. Kalidas, and N. Chandra. targetTB: A target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst Biol*, 2(1):109, 2008
- K. Raman, R. Vashisht, and N. Chandra. Strategies for efficient disruption of metabolism in *Mycobacterium tuberculosis* from network analysis. *Mol Biosyst*, 5:1740–1751, 2009
- K. Raman, A. G. Bhat, and N. Chandra. A systems perspective of hostpathogen interactions: Predicting disease outcome in tuberculosis. *Mol Biosyst*, 6:516–530, 2010
- K. Raman, Y. Kalidas, and N. Chandra. Model-driven drug discovery: Principles and practices, *Biological database modeling*, pages 163–188. Artech House, New York, 2007
- J. C. J. Ray and D. E. Kirschner. Requirement for multiple activation signals by anti-inflammatory feedback in macrophages. *J Theor Biol*, 241(2):276–294, 2006
- J. C. J. Ray, J. Wang, J. Chan, and D. E. Kirschner. The timing of TNF and IFN- $\gamma$  signaling affects macrophage activation strategies during *Mycobacterium tuberculosis* infection. *J Theor Biol*, 252(1):24–38, 2008
- J. L. Reed, I. Famili, I. Thiele, and B. Ø. Palsson. Towards multidimensional genome annotation. *Nat Rev Genet*, 7(2):130–141, 2006a
- J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, D. D. Herring, O. T. Bui, E. M. Knight, S. S. Fong, and B. Ø. Palsson. Systems approach to refining genome annotation. *Proc Natl Acad Sci USA*, 103(46):17480–17484, Nov 2006b
- C. M. Sassetti, D. M. Boyd, and E. J. Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol*, 48(1):77–84, 2003
- D. Schnappinger, S. Ehrt, M. I. Voskuil, Y. Liu, J. A. Mangan, I. M. Monahan, G. Dolganov, B. Efron, P. D. Butcher, C. Nathan, and G. K. Schoolnik. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. *J Exp Med*, 198(5):693–704, 2003
- J. L. Segovia-Juarez, S. Ganguli, and D. E. Kirschner. Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *J Theor Biol*, 231(3):357–376, 2004
- I. Smith. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*, 16(3):463–496, 2003

- M. Strong, T. G. Graeber, M. Beeby, M. Pellegrini, M. J. Thompson, T. O. Yeates, and D. Eisenberg. Visualization and interpretation of protein networks in *Mycobacterium tuberculosis* based on hierarchical clustering of genome-wide functional linkage maps. *Nucleic Acids Res*, 31(24):7099–7109, 2003
- D. Sud, C. Bigbee, J. L. Flynn, and D. E. Kirschner. Contribution of CD8<sup>+</sup> T cells to control of *Mycobacterium tuberculosis* infection. *J Immunol*, 176(7):4296–4314, 2006
- K. Takayama, C. Wang, and G. S. Besra. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin Microbiol Rev*, 18:81–101, 2005
- J. Thakar, M. Piliore, G. Kirimanjeswara, E. T. Harvill, and R. Albert. Modeling systems-level regulation of host immune responses. *PLoS Comput Biol*, 3(6):e109, 2007
- T. Thomas. Boolean formalization of genetic control circuits. *J Theor Biol*, 42(3):563–585, 1973
- K. D. Verkhedkar, K. Raman, N. Chandra, and S. Vishveshwara. Metabolome based reaction graphs of *M. tuberculosis* and *M. leprae*: A comparative network analysis. *PLoS One*, 2(9): e881, 2007
- P. K. Vinod, B. Konkimalla, and N. Chandra. *In-silico* pharmacodynamics: Correlation of adverse effects of H2-antihistamines with histamine N-methyl transferase binding potential. *Appl Bioinform*, 5(3):141–150, 2006
- C. Von Mering, L. J. Jensen, M. Kuhn, S. Chaffron, T. Doerks, B. Krüger, B. Snel, and P. Bork. STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res*, 35(Database issue):358–362, 2007
- J. S. Waddell and P. D. Butcher. Microarray analysis of whole genome expression of intracellular *Mycobacterium tuberculosis*. *Curr Mol Med*, 7(3):287–296, 2007
- S. J. Waddell, R. A. Stabler, K. Laing, L. Kremer, R. C. Reynolds, and G. S. Besra. The use of microarray analysis to determine the gene expression profiles of *Mycobacterium tuberculosis* in response to anti-bacterial compounds. *Tuberculosis (Edinb)*, 84(3–4):263–274, 2004
- S. J. Waddell, P. D. Butcher, and N. G. Stoker. Rna profiling in host-pathogen interactions. *Curr Opin Microbiol*, 10(3):297–302, 2007
- S. J. Waddell, K. Laing, C. Senner, and P. D. Butcher. Microarray analysis of defined *Mycobacterium tuberculosis* populations using rna amplification strategies. *BMC Genomics*, 9:94, 2008
- J. E. Wigginton and D. E. Kirschner. A model to predict cell-mediated immune regulatory mechanisms during human infection with *Mycobacterium tuberculosis*. *J Immunol*, 166(3): 1951–1967, 2001
- World Health Organisation. *Global tuberculosis control: Surveillance, planning, financing: WHO report 2008*. World Health Organisation, 2008 ISBN 978-9241563543
- K. Yeturu and N. Chandra. PocketMatch: A new algorithm to compare binding sites in protein structures. *BMC Bioinform*, 9:543, 2008
- D. Young, J. Stark, and D. E. Kirschner. Systems biology of persistent infection: Tuberculosis as a case study. *Nat Rev Microbiol*, 6(7):520–528, 2008

# Chapter 6

## Qualitative Analysis of Genetic Regulatory Networks in Bacteria

Valentina Baldazzi, Pedro T. Monteiro, Michel Page, Delphine Ropers, Johannes Geiselmann, and Hidde de Jong

### 6.1 Introduction

The functions of living organisms are controlled on the molecular level by networks of biochemical reactions involving genes, mRNAs, proteins, metabolites, and signaling molecules. The elucidation of the structure of these networks has much progressed thanks to decades of work in genetics, molecular biology, and biochemistry, including the development of high-throughput experimental techniques. Most of the time, however, it is not well understood how the dynamics of the networks emerge from the reactions between the variety of its individual components. This has called forth an increasing interest in the mathematical modeling of complex cellular processes, in the context of a broader movement called systems biology (Bettenbrock et al. 2005; Chen et al. 2004; Klipp et al. 2005; Leloup and Goldbeter 2003; Schoeberl et al. 2002).

In theory, it is possible to write down kinetic models of biochemical networks, and study these by means of classical analysis and simulation tools. In practice, this is not easy to achieve though, as the values of kinetic parameters are often only constrained to within a range spanning several orders of magnitude for most systems of biological interest. Moreover, the models consist of a large number of variables, are strongly non-linear, and include different timescales, which makes them difficult to handle both mathematically and computationally. This has motivated the use of approximations reducing the size and complexity of the models. Various approximations have been proposed in the literature, tailored to typical response functions and timescale hierarchies found in genetic or metabolic regulation (de Jong and Ropers 2006; Heijnen 2005; Heinrich and Schuster 1996; Millat et al. 2007; Okino and Mavrouniotis 1998; Papin et al. 2004; Pecou 2005; Roussel and Fraser 2001; Savageau 2001; Thomas and Kaufman 2001). The approximations typically reduce the dimension of the state and parameter space, and they simplify the mathematical form of the equations.

---

V. Baldazzi (✉)  
INRIA Grenoble – Rhône-Alpes, France  
e-mail: [Valentina.Baldazzi@inria.fr](mailto:Valentina.Baldazzi@inria.fr)

In this chapter, we discuss a class of so-called *piecewise-linear (PL) models* of genetic regulatory networks, based on the use of step-function approximations of the sigmoidal response functions involved in gene regulation. The PL models, originally introduced by Glass and Kauffman (1973) provide a coarse-grained picture of the dynamics of genetic regulatory networks. They associate a protein concentration variable to each of the genes in the network, and capture the switch-like character of gene regulation by means of step functions that change their value at a threshold concentration of the proteins. The advantage of using PL models is that the qualitative dynamics of the high-dimensional systems are relatively simple to analyze, using an ordering of parameters rather than exact numerical values (Batt et al. 2008; de Jong et al. 2004b). This makes the PL models a valuable tool for the analysis of genetic regulatory networks, as demonstrated by several examples in bacteria and higher organisms (Chaves et al. 2006; de Jong et al. 2004a; Halász et al. 2007; Ropers et al. 2006; Sepulchre et al. 2007; Usseglio Viretta and Fussenegger 2004).

We will introduce the PL models in the context of the network of global regulators controlling the carbon starvation response in the enterobacterium *Escherichia coli*. In order to survive, *E. coli* cells constantly have to adapt their functioning to the availability of carbon sources, essential for growth. The adaptation involves multiple levels of regulation, from metabolic fluxes and enzyme activity to gene regulation (Gutierrez-Ríos et al. 2007; Hardiman et al. 2007; Kremling et al. 2009). In this chapter, we focus in particular on the role of the global regulators of transcription, such as CRP, Fis, DNA supercoiling, and RpoS. These global regulators form the backbone of the network coordinating the long-term response of *E. coli* cells to starvation conditions.

In Sect. 6.2 we briefly review the carbon starvation response in *E. coli* and the role of the global regulators. Section 6.3 describes how we can systematically reduce a classical kinetic model of the network of global regulators to a PL model. The different mathematical and computational techniques available for the analysis of PL models are discussed in Sects. 6.4 and 6.5, and illustrated on the *E. coli* model. A particularly interesting property of the PL models is that they allow the parameter space to be partitioned into regions with the same qualitative dynamics, using sets of simple inequality constraints between parameters. This property is exploited in Sect. 6.6 for the analysis of incompletely specified models. Section 6.7 discusses the strengths and limitations of the PL models, and their relation to other qualitative models, such as Boolean networks.

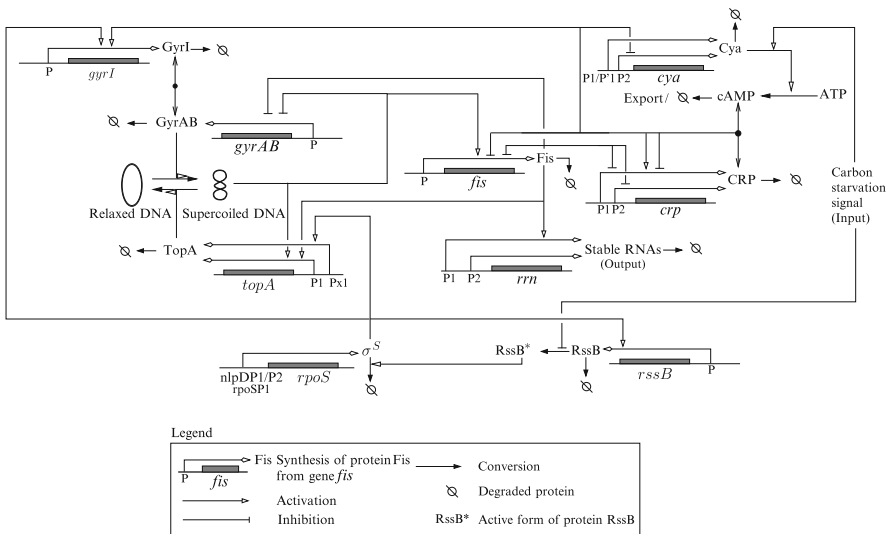
## 6.2 Carbon Starvation in *E. coli*

Under favorable environmental conditions, bacterial cells quickly grow and divide, leading to an exponential increase of their biomass, called *exponential phase*. Upon a variety of stress conditions, like the depletion of carbon sources, the bacteria abandon exponential phase and enter a state in which cells stop dividing, capitalizing upon the few available resources to maintain the basic metabolic functions

necessary for survival. This so-called *stationary phase* is rapidly reversed and fast growth restored once the environmental conditions become favorable again (Huisman et al. 1996).

Glucose is the preferred carbon source of *E. coli*. The adaptation of the bacteria to the depletion of glucose from the growth medium is under the control of a large and complex network of biochemical reactions involving genes, mRNAs, proteins, metabolites, and signaling molecules. The role of the metabolic and signaling networks in the adaptation of *E. coli* to carbon source starvation have been extensively studied (e.g., Bettenbrock et al. 2006; Chassagnole et al. 2002; Rohwer et al. 2000), but much less has been done at the level of gene expression. In particular, it is not well understood how the backbone of the network, formed by the global regulators of transcription, coordinates the cascades of molecular events driving the growth arrest of *E. coli* cells starved for glucose. These transcription factors respond directly or indirectly to glucose depletion, by controlling in a combinatorial fashion the expression of a large number of genes involved in cellular adaptation and survival. In addition, they control each other's expression, thus giving rise to a complex regulatory network.

Figure 6.1 shows the network of key global regulators involved in the control of the carbon starvation response. It includes well-known pleiotropic transcription regulators, like the histone-like protein Fis, the catabolic repressor cAMP·CRP (resulting from the expression of genes *crp* and *cya*, and the activation of Cya by carbon depletion), and the general stress response factor RpoS or  $\sigma^S$  (whose stability is regulated by RssB). Changes in DNA topology and its dependence on the relative



**Fig. 6.1** Network of global regulators involved in the carbon starvation response in *E. coli* (Ropers et al., 2011, 2006). The graphical conventions (Kohn 2001) are explained in the legend

expression level of the genes *gyrA*, *gyrB*, *gyrI*, and *topA* are also considered, as the three-dimensional structure of DNA modulates the transcription of a large number of genes. Finally, stable RNAs expressed from the *rrn* operons are considered as their amount provides a reliable indicator of the growth rate of the cell, being high during an exponential phase and low during a stationary phase.

The synthesis and degradation of proteins and stable RNAs, and their regulation by the global regulators, are central processes in the system of Fig. 6.1. However, the network also involves other types of biochemical reactions, such as the formation of protein complexes (GyrAB-GyrI), the modification of proteins by small molecules (cAMP-CRP), and enzymatic reactions (the synthesis of cAMP by Cya).

### 6.3 Modeling and Model Reduction

Due to the size and complexity of the network, the dynamics of the carbon starvation response is difficult to predict intuitively and a mathematical model may be a useful tool to clarify the global behavior of the system. Modeling may also allow the formulation of hypotheses about missing components of the system, opening the way to further experimental investigations.

A wide variety of modeling formalisms are available to describe networks of biochemical reactions. The most common approach is based on *ordinary differential equations (ODEs)* and describes the rate of change of the concentrations of proteins, RNAs, metabolites, and other molecular species in the network. ODE models have a solid foundation in the kinetic theory of biochemical reactions (Cornish-Bowden 1995; Heinrich and Schuster 1996), but their application requires knowledge of the precise molecular mechanisms involved and quantitative information on parameter values. For many systems, like the network of global regulators in *E. coli*, this level of knowledge is not available. We therefore propose a model-reduction strategy based on *quasi-steady-state (QSS)* and *PL* approximations, in order to obtain models that are easier to handle mathematically and computationally.

We start from a detailed nonlinear ODE model that we build following standard approaches from biochemistry. Figure 6.1a–b shows a small part of the ODE model, concerning the activation of the transcription factor CRP, which we will use as example network throughout the chapter (see Ropers et al. (2011) for the complete model). Depending on their type, the reactions are assumed to follow mass-action, Michaelis–Menten, or Hill kinetics. The resulting ODE system is highly nonlinear and depends on a large number of parameters, whose values that are mostly unknown within a range of several orders of magnitude.

A first reduction step is motivated by the fact that the processes described by the model occur on widely differing timescales. Even in the absence of precise information on the parameter values, we can to a first approximation distinguish two different timescales. The first is a fast timescale, corresponding to reactions like the formation of the complexes cAMP-CRP and the metabolic reactions responsible for the synthesis, export, and degradation of cAMP. The second is a

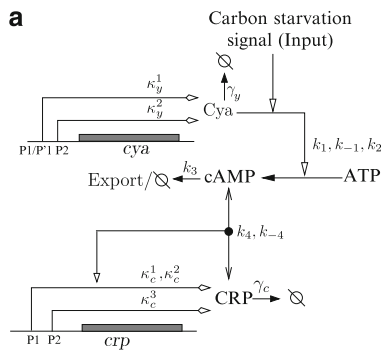
slow timescale corresponding to the synthesis and degradation of the proteins Cya and CRP. Based on timescale separation, the original model can be rewritten into two distinct subsystems, corresponding to slow processes (protein synthesis and degradation) and fast processes (complex formation and enzymatic reactions). The fast and the slow processes are described by so-called fast and slow variables, respectively.

To reduce the size and complexity of the nonlinear model, we apply the QSS assumption (Heinrich and Schuster 1996). The QSS assumption is based on the hypothesis that the fast variables instantaneously adapt to changes in the slow variables. This means that, after an initial transient, the dynamics of the fast system can be well approximated by an algebraic function of the slow variables. The algebraic function replaces the ODEs for the fast subsystem. Figure 6.1c compares the temporal evolution of the concentration of cAMP-CRP complex that is predicted by the nonlinear and QSS models, showing a good agreement between the two solutions.

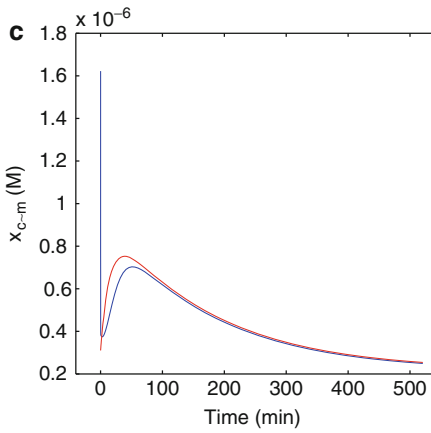
The QSS model thus obtained (Fig. 6.1d) is still nonlinear, but its dimension is reduced to the slow variables Cya and CRP only, whose dynamics completely defines the changes in the amount of cAMP-CRP complex (the fast variable).

As a further approximation we simplify the mathematical form of the nonlinearities of the system. This can be obtained by using PL differential equations. This simplification is motivated by the fact that at the slow timescale of interest the regulation of gene expression gives rise to multivariate sigmoidal response functions, which are conveniently approximated by algebraic expressions (e.g., sums or products) of step functions. Following the QSS approximation, the control exerted by cAMP-CRP on the synthesis of CRP protein can be rewritten as a function of Cya and CRP concentrations (Fig. 6.1e). Gene expression appears to be maximal when both Cya and CRP concentration are at high levels, within their physiological range, and rapidly reduces for smaller concentrations of these proteins. Such a behavior can be well approximated by a product of two-step functions,  $s^+(x_y, \theta_y^1)$  and  $s^+(x_c, \theta_c^1)$ , where  $x_y$  and  $x_c$  represent the concentrations of Cya and CRP proteins, respectively, and  $\theta_y^1$  and  $\theta_c^1$  the threshold concentrations of these proteins. The step-function expressions are equivalent to logical functions, and account for the combinatorial control of gene expression by regulatory proteins (Glass and Kauffman 1973; Mestl et al. 1995; Thomas and d'Ari 1990). The PL model which includes these step functions is shown in Fig. 6.1f. It states that, in the presence of a carbon starvation signal, the gene *crp* will be expressed at its maximal rate if Cya and CRP protein concentrations are simultaneously larger than their threshold value, respectively  $\theta_y^1$  and  $\theta_c^1$ . When this condition is not fulfilled, the synthesis of CRP protein is simply reduced to its basal rate.

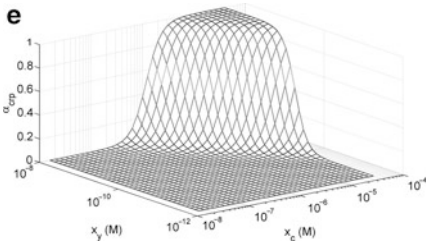
The above reduction strategy has been applied to the whole carbon starvation network in Fig. 6.1, resulting in a PL model with 9 variables and 50 parameters (Ropers et al. 2011). The quality of the QSS and PL approximations has been systematically tested, using an ensemble approach and appropriate distance and correlation measures to compare the numerical solutions of the different models at each reduction step (Ropers et al. 2011). The results show that, in comparison with conventional

**b**

$$\begin{aligned}\dot{x}_{y\sim} &= \kappa_y^1 + \kappa_y^2 h^+(x_{c\sim m}, \theta_{c\sim m}^2, m_{c\sim m}) - \gamma_y x_{y\sim} \\ &\quad + (k_{-1} + k_2 h^-(u_s, \theta_s, m_s)) x_{y\sim p} \\ \dot{x}_{c\sim} &= \kappa_c^1 + \kappa_c^2 h^+(x_{c\sim m}, \theta_{c\sim m}^1, m_{c\sim m}) - \gamma_c x_{c\sim} \\ &\quad + k_{-4} x_{c\sim m} - k_4 x_{c\sim} x_{m\sim} \\ \dot{x}_{y\sim p} &= k_1 x_{y\sim} x_p - (k_{-1} + k_2 h^+(u_s, \theta_s, m_s) + \gamma_y) x_{y\sim p} \\ \dot{x}_{c\sim m} &= k_4 x_{c\sim} x_{m\sim} - (k_{-4} + \gamma_c) x_{c\sim m} \\ \dot{x}_{m\sim} &= k_2 h^+(u_s, \theta_s, m_s) x_{y\sim p} + k_{-4} x_{c\sim m} - k_3 x_{m\sim} \\ &\quad - k_4 x_{c\sim} x_{m\sim}\end{aligned}$$

**d**

$$\begin{aligned}\dot{x}_y &= \kappa_y^1 + \kappa_y^2 - \gamma_y x_y \\ \dot{x}_c &= \kappa_c^1 + \kappa_c^2 h^+(x_{c\sim m}, \theta_{c\sim m}^1, m_{c\sim m}) - \gamma_c x_c \\ x_{c\sim m} &= \frac{h^+(u_s, \theta_s, m_s) x_c x_y}{K_4 K_3 + h^+(u_s, \theta_s, m_s) x_y} \\ K_3 &= \frac{k_3}{k_2} \\ K_4 &= \frac{k_{-4}}{k_4}\end{aligned}$$

**f**

$$\begin{aligned}\dot{x}_y &= \kappa_y^1 + \kappa_y^2 - \gamma_y x_y \\ \dot{x}_c &= \kappa_c^1 + \kappa_c^2 s^+(x_c, \theta_c^1) s^+(x_y, \theta_y^1) s^+(u_s, \theta_s) - \gamma_c x_c\end{aligned}$$

nonlinear models, the PL approximations generally preserve the dynamics of the carbon starvation response network, encouraging the use of PL models in situations where the reference timescale is that of protein synthesis and degradation.

## 6.4 Qualitative Analysis of Dynamics

The dynamical properties of PL models have been the subject of active research for more than three decades (Batt et al. 2007; Edwards 2000; Ghosh and Tomlin 2004; Glass and Kauffman 1973; Gouzé and Sari 2002; Mestl et al. 1995; Plahte and Kjógglum 2005). The models have favourable mathematical properties due to the use of step functions in the right-hand side of the differential equations. The thresholds of the concentration variables define a subdivision of the state space, the set of possible states of the system, into hyperrectangular regions.

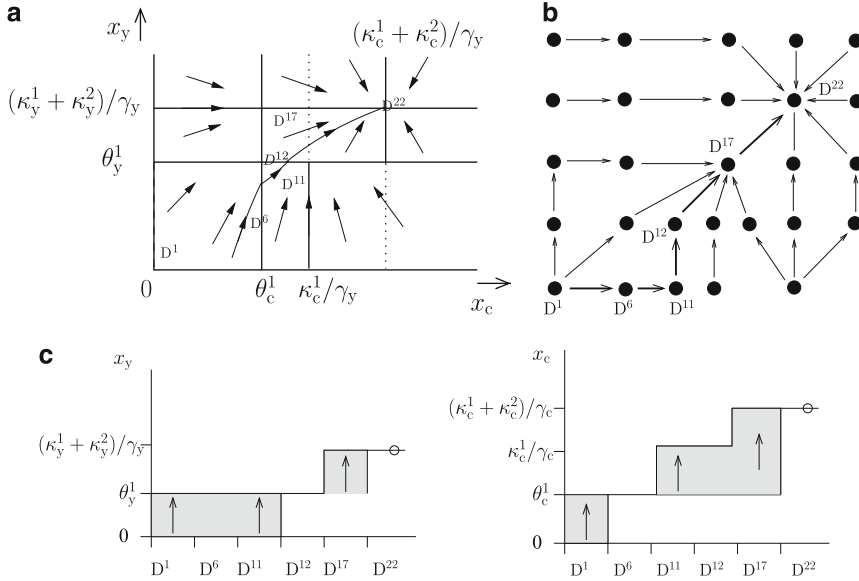
In every region not located on a threshold plane, the step functions evaluate to 0 or 1, and the PL model reduces to an analytically solvable system of differential equations. For instance, in the region  $0 < x_y < \theta_y^1$  and  $0 < x_c < \theta_c^1$  (named  $D^1$  in Fig. 6.2), our example model (Fig. 6.1f) simply reduces to

$$\begin{aligned}\dot{x}_y &= k_y^1 + k_y^2 - \gamma_y x_y \\ \dot{x}_c &= k_c^1 - \gamma_c x_c,\end{aligned}\tag{6.1}$$

as both  $s^+(x_c, \theta_c^1)$  and  $s^+(x_y, \theta_y^1)$  evaluate to 1 in  $D^1$ . It can be shown that all solutions of (6.1) monotonically converge towards the point  $((k_y^1 + k_y^2)/\gamma_y, k_c^1/\gamma_c)$ . The point  $((k_y^1 + k_y^2)/\gamma_y, k_c^1/\gamma_c)$  is called a focal point of the system. Setting further  $(k_y^1 + k_y^2)/\gamma_y > \theta_y^1$  and  $k_c^1/\gamma_c > \theta_c^1$ , as explained below, it immediately follows that Cya and CRP concentrations increase everywhere in region  $D^1$ . The above example holds more generally: in every region  $D$  the original PL model simplifies

---

**Fig. 6.1** *Nonlinear ODE model that we build following standard approaches from biochemistry. (a) Activation network. (b) Detailed ODE model for the activation network.  $x_{y\sim}, x_{c\sim}, x_{y\sim p}, x_{c\sim m}$ , and  $x_{m\sim}$  denote the concentrations of free Cya, free CRP, Cya·ATP, cAMP·CRP, and free cAMP, respectively, while  $u_c$  denotes the external glucose concentration. The total concentrations of Cya, CRP, and cAMP are referred to as  $x_y, x_c$ , and  $x_m$ , respectively.  $h^+$  denotes a positive Hill function:  $h^+(x, \theta, m) = \frac{x^m}{x^m + \theta^m}$ . (c) nonlinear and QSS solutions. The blue curve represents a solution for the concentration variable  $x_{c\sim m}$  in the nonlinear model. The red curve is the corresponding solution for  $x_{c\sim m}$  in the QSS model. After an initial transient the nonlinear solution rapidly relaxes to the QSS solution. (d) QSS model for the activation network. The model approximates the nonlinear model by coupling the fast variable  $x_{c\sim m}$  to the slow variables  $x_y$  and  $x_c$ . (e) Plot of  $h^+(x_{c\sim m}, \theta_{c\sim m}, m_{c\sim m})$  as a function of  $x_y$  and  $x_c$ . The sigmoidal surface is approximated by the product of step functions  $s^+(x_y, \theta_y) s^+(x_c, \theta_c) s^+(u_c, \theta_s)$ , with  $\theta_c, \theta_y$ , and  $\theta_s$  threshold values.  $s^+$  denotes a positive-step function:  $s^+(x, \theta) = 1$  if  $x > \theta$ , and 0 if  $x < \theta$ . (f) PL model for the activation network with the step = function approximation of  $h^+(x_{c\sim m}, \theta_{c\sim m}, m_{c\sim m})$*



**Fig. 6.2** *State-space and state transition for PL model.* (a) State-space diagram for the PL model of the subsystem controlling CRP activation (Fig. 6.1). The parameter values are assumed to satisfy the following inequalities:  $\kappa_c^1/\gamma_c > \theta_c^1$ ,  $(\kappa_c^1 + \kappa_c^2)/\gamma_c > \theta_c^1$ , and  $(\kappa_y^1 + \kappa_y^2)/\gamma_y > \theta_y^1$ . The arrows indicate the direction of the vector field in each of the regions of the state space, while the solid line represents a solution trajectory evolving towards a stable steady state in the region  $D^{22}$ . (b) State transition graph representing the qualitative dynamics of the PL model. The state  $D^1$  corresponds to the region with the same name in part (a) of the figure, and satisfies the atomic properties  $0 \leq x_c < \theta_c$ ,  $0 \leq x_y < \theta_y$ ,  $dx_c/dt > 0$ , and  $dx_y/dt > 0$ . The labeled path ( $D^1$ ,  $D^6$ ,  $D^{11}$ ,  $D^{12}$ ,  $D^{17}$ ,  $D^{22}$ ) in the graph corresponds to the solution trajectory in (a). (c) Visualization of the concentration bounds and signs of derivatives for each of the variables along the path in (b). It shows that temporal-logic property (6.2) in the main text is satisfied by the graph. That is, the graph contains a path in which the CRP concentration is increasing before becoming steady

into set of linear differential equations such that the systems locally behaves in a qualitatively homogeneous way, all solutions monotonically converging towards a focal point given by the ratio of (sum of) synthesis rate constants and (sum of) degradation rate constants. These results can be generalized to the case of regions located on a threshold plane (Batt et al. 2007; Gouzé and Sari 2002). Figure 6.2a shows the subdivision of the state space for the PL model of the simple network controlling the activation of CRP, as well as the dynamics in each of the regions.

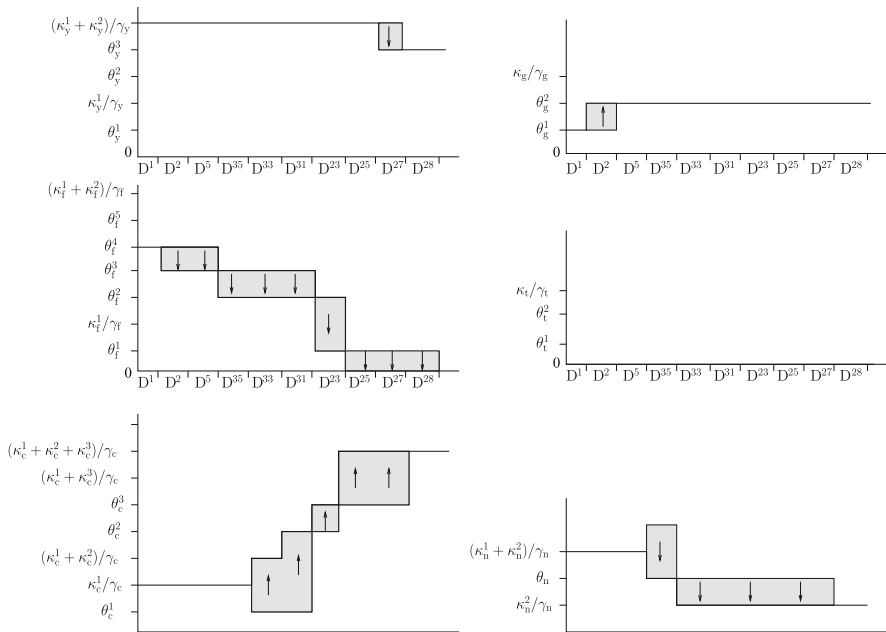
The fact that the system behaves in a qualitatively homogeneous way in every region motivates a discrete representation of the dynamics of the PL models, by means of a so-called state transition graph. The states in the graph correspond to the regions in the state space, while the transitions arise from solutions that enter one region from another. Each transitions thus corresponds to a discrete event, namely the crossing of a threshold by one or more concentration variables, possibly entailing a change in the derivative (trend) of these variables.

The generation of the state transition graph from a PL model has been implemented in the computer tool Genetic Network Analyzer (GNA) (Batt et al. 2005; de Jong and Page 2008; de Jong et al. 2003), explicitly developed for the simulation and the analysis of PL models. The state transition graph associated with the example network is shown in Fig. 6.2b. The paths in the graphs denote sequences of qualitative events, notably threshold crossings of the variables and changes in the sign of the derivatives, as illustrated in panel (c) of Fig. 6.2.

Interestingly, it can be shown that the state transition graph, and thus the qualitative dynamics of the system, are completely determined by inequality constraints defining the ordering between the threshold parameters  $\theta_i$  of a variable  $x_i$  and the values of the focal points for that variable. The definition of the inequality constraints between parameters can generally be inferred from available data in the experimental literature or by intuitive reasoning, even in the absence of quantitative information on parameter values. For example, the position of the focal point concentration  $(\kappa_y^1 + \kappa_y^2)/\gamma_y$  for protein Cya can be deduced just noticing that the ratio  $(\kappa_y^1 + \kappa_y^2)/\gamma_y$  defines the maximum steady-state concentration that Cya can reach. In order for Cya to have a regulatory effect on CRP synthesis, it is thus natural to assume that  $(\kappa_y^1 + \kappa_y^2)/\gamma_y > \theta_y^1$ , otherwise the regulation of *crp* expression would never be functional. GNA exploits the inequality constraints to symbolically compute the attractors of the model and the states that are reachable from given initial conditions for the concentration variables. This process is called qualitative simulation.

We carry out a qualitative simulation of the carbon starvation network in Fig. 6.1. Given the PL model of the *E.coli* network obtained in the previous section, we define a set of inequality constraints on the parameters, as illustrated in Fig. 6.2 for the CRP activation subnetwork. The full specification of the PL model requires more than 50 inequality constraints (Ropers et al. 2011). The PL model has been used to address the question whether the interactions between the global regulators in Fig. 6.1 are sufficient to explain the growth adaptation of *E. coli*. We analyze the attractors of the system and simulate the response of *E. coli* to a depletion or sudden availability of glucose.

Attractor analysis identifies the presence of two stable steady states. The first steady state, characterized by the presence of the carbon starvation signal and by a low level of stable RNAs, corresponds to stationary-phase conditions, whereas the second steady state is representative of exponential phase, with a high level of stable RNAs and no carbon starvation. Depending on the presence or absence of glucose in the growth medium, the bacteria reach one of the attractors and are in exponential or stationary phase, respectively. In order to better investigate the dynamics of the transitions between growth phases, we simulate the qualitative behavior of the network, starting from the steady state corresponding to exponential phase (stationary phase) and perturbing the system by switching on (off) the carbon starvation signal. An example of a path in the state transition graph produced by the qualitative simulation of the entry into stationary phase is shown in Fig. 6.3. The path traces how the concentrations of the global regulators Fis and CRP evolve during the transition from exponential to stationary phase. In response to a lack of glucose, the first key



**Fig. 6.3** Path in the state transition graph. Path in the state transition graph produced by a qualitative simulation of the entry into stationary phase following carbon depletion, using the PL model of the network in Fig. 6.1. The evolution of Cya ( $x_y$ ), Fis ( $x_f$ ), Ctp ( $x_c$ ), GyrAB ( $x_g$ ), TopA ( $x_t$ ) and the stable RNAs ( $x_n$ ) is shown

event is a decrease in the concentration of Fis protein, followed by an progressive increase of the level of CRP, strengthening the negative effect on *fis* expression. The decrease in Fis levels ultimately leads to a downregulation of the level of stable RNAs, witnessing the growth arrest as the cells enter stationary phase.

The predicted evolution of the Fis concentration is in agreement with experimental data showing a 50-fold decrease in protein levels when *E. coli* cells enter stationary phase (Ali Azam et al. 1999; Ball et al. 1992; Pratt et al. 1997). Unfortunately, similar measurement are not available for CRP, but some experimental observations tend to confirm the model predictions: a low protein concentration has been measured in the presence of glucose, whereas CRP is shown to accumulate when glucose is absent (Ishizuka et al. 1994).

The question can be raised how general the above conclusions are. The simulation results shown in Fig. 6.3 represent just one of the possible qualitative behaviors of the system starting from the initial conditions, among a large number of paths in the state-transition graph. Indeed, the entire state space for the PL model of the network in Fig. 6.1 consists of the order of  $10^{10}$  states whereas the subset reachable from a particular growth phase of the bacteria contains approximately  $10^3$  states. The size of the graph makes a general statement on the system dynamics difficult to obtain by manual inspection of the individual simulations.

## 6.5 Formal Verification of Network Properties

The formal verification field provides powerful methods to deal with the analysis of large models of cellular interaction networks by specifying dynamical properties of interest, as statements in a formal language called temporal logic (Antoniotti et al. 2003; Batt et al. 2005; Bernot et al. 2004; Calder et al. 2005; Chabrier-Rivier et al. 2004; Fisher et al. 2007). Efficient, so-called model-checking, algorithms exist to determine whether these statements are satisfied by the model, without the need of explicitly checking all the paths in the graph (Clarke et al. 1999).

Temporal logic queries are meant to capture patterns in the temporal evolution of the system dynamics, like the relation and the ordering between qualitative events (e.g., increasing/decreasing of a protein concentration) or the reachability of an attractor of the system (a steady state or a limit cycle). For example, given the model for CRP activation, we can ask whether there exists a path in which the concentration of CRP increases before becoming steady. Such a property can be specified in temporal logic by the following formula

$$EF(dx_c/dt > 0 \wedge EF(dx_c/dt = 0)) \quad (6.2)$$

that is satisfied by the state transition graph in Fig. 6.2, as witnessed by the path in panel (c). The definition of specific queries can be difficult for nonexpert users though, due to the existence of a variety of temporal logic operators that can be applied alone or in combination, with sometimes subtle differences in the meaning of the resulting formulas. Alternatively, high-level query templates can be defined that link the intuitive description of biological properties to temporal logic (Monteiro et al. 2008).

We apply model-checking tools to the analysis of the state transition graphs produced by qualitative simulation of the carbon starvation response in *E. coli*, investigating the role of the mutual inhibition between the genes *fis* and *crp*. When found in isolation, a mutual-inhibition motif has been shown to lead to bistability (Gardner et al. 2000): in particular, it excludes the simultaneous presence of high concentrations of the proteins CRP and Fis, as shown in the simulation in Fig. 6.3. The question can be asked if this is always true, i.e., if this motif maintains its functionality when embedded in a large network like the one in Fig. 6.1. If this is the case, it should be impossible for the concentrations of Fis and CRP to be simultaneously high or low at steady state. Reformulate into a statement in temporal logic, in particular the logic CTL, following (Monteiro et al. 2008), this property reads:

$$\begin{aligned} &\neg EF (high(x_c) \wedge high(x_f) \wedge isSteadyState) \\ &\wedge \neg EF (low(x_c) \wedge low(x_f) \wedge isSteadyState). \end{aligned} \quad (6.3)$$

The property states that, starting from the initial conditions, there does not exist a future state ( $\neg EF$ ) where  $x_c$  and  $x_f$  simultaneously have a high value and the

system is in steady state ( $high(x_c) \wedge high(x_f) \wedge isSteadyState$ ), and similarly, there does not exist a future state where  $x_c$  and  $x_f$  simultaneously have a low value and the system is in steady state ( $low(x_c) \wedge low(x_f) \wedge isSteadyState$ ). The predicates *high* and *low* are defined in terms of inequality constraints. For instance,  $high(x_f)$  can be defined as  $x_f > \theta_f^4$  for a high threshold  $\theta_f^4$  of Fis in the PL model.

Formal verification using the open-source model-checking tool NuSMV (Cimatti et al. 2002), in combination with GNA, reveals that the property holds true for all possible paths in the state-transition graph. The positive loop motif involving *fis* and *crp* thus remains functional inside the large network of Fig. 6.1, with the consequent rearrangements of gene expression levels following the transition from a state with high Fis and low CRP concentrations (characteristic of the exponential growth phase) to a situation with low Fis and high CRP (typical of the stationary growth phase).

The carbon starvation response network also contains a negative feedback loop, involving the genes *gyrAB*, *topA*, and *fis* (Fig. 6.1). GyrAB and TopA are responsible for the control of the intracellular level of DNA supercoiling. GyrAB is a gyrase protein which supercoils the DNA structure, whereas the topoisomerase TopA relaxes it. An increase of the DNA supercoiling level stimulates expression of Fis, which in turn decreases the supercoiling level, by stimulating *topA* expression and inhibiting *gyrAB* expression. This negative loop allows the bacteria to rapidly adjust protein concentrations and resume growth once nutrients become available again. Qualitative simulations have shown the emergence of damped oscillations of Fis and GyrAB concentrations in response to a sudden carbon upshift. Is this property bound to occur following a carbon upshift? We used formal verification methods to check whether the carbon upshift is a sufficient condition for the occurrence of damped oscillations (Monteiro et al. 2008), i.e.:

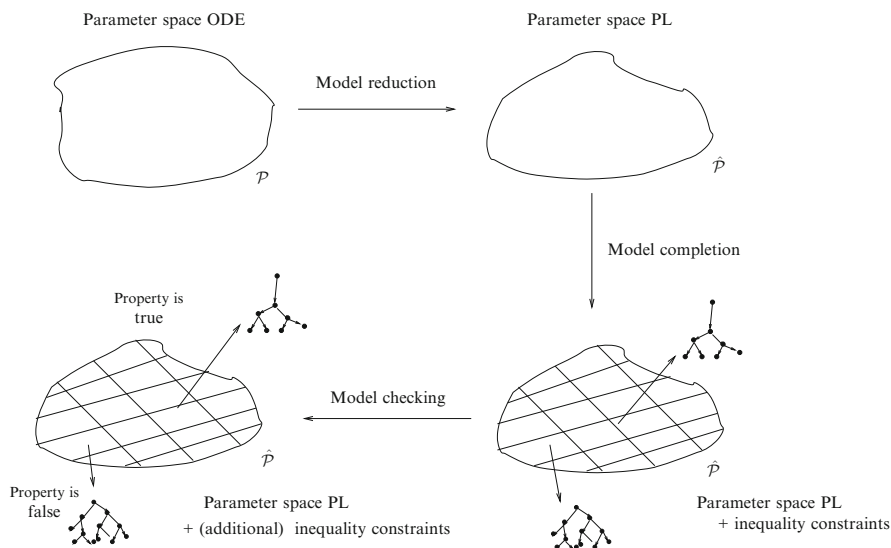
$$AG ((u_s < \theta_s) \Rightarrow AF (isOscillatoryState(x_f, x_g))). \quad (6.4)$$

The predicate  $isOscillatoryState(x_f, x_g)$  denotes that a state is part of a cycle in the graph in which the concentrations of Fis and GyrAB oscillate.

The model checker returned true for the query: the model thus predicts that cells necessarily resume growth through damped oscillations after a carbon upshift. However, no experimental data are currently available to confirm or disconfirm this prediction.

## 6.6 Model Completion

As explained in Sect. 6.4, the state-transition graph defining the qualitative dynamics is completely determined by the inequality constraints on the parameters. Each set of inequality constraints defines a region in the parameter space in which the system has the same qualitative dynamics (Batt et al. 2008). In case it is not possible to completely specify the ordering of the parameters, formal verification can



**Fig. 6.4** Overall scheme for model completion by means of formal verification tools. A nonlinear ODE model with parameter space  $\mathcal{P}$  is reduced to a PL model with parameter space  $\hat{\mathcal{P}}$ . This parameter space can be partitioned into regions with the same qualitative dynamics, represented by the state transition graph, using inequality constraints on the parameter values. By testing properties that have to be satisfied by the model, certain regions of the parameter space can be eliminated, thus further constraining the parameter values

provide a method to discriminate between alternative hypotheses (Bernot et al. 2004). Different sets of inequality constraints can be enumerated and the dynamics of the corresponding PL models tested against known biological properties by means of model checking. Sets of inequality constraints that are proven to be inconsistent with the properties are rejected and further analysis can be performed on the biologically-meaningful models left. This model completion approach allows the exhaustive exploration of the parameter space of the PL models (Fig. 6.4).

As an example, we investigate the role played by the general stress factor RpoS in the control of the DNA supercoiling level and in the growth adaptation of *E. coli* cells following carbon depletion. The global regulatory network in Fig. 6.1 extends a previously published PL model (Ropers et al. 2006), which failed to predict the observed decrease of the DNA supercoiling level during the transition from the exponential to the stationary growth phase. We therefore refined the description of the control of DNA supercoiling level by including in the model the general stress factor RpoS and its regulators, since a *rpoS* deletion mutant has been shown deficient in the regulation of plasmid topology in stationary phase (Reyes-Dominguez et al. 2003). How does RpoS control the DNA supercoiling level during the adaptation of *E. coli* cells to carbon depletion? And is the entry into stationary phase always preceded by an accumulation of RpoS?

Model completion was carried out by taking a partially specified PL model of the network in Fig. 6.1. The model does not constrain the ordering of the threshold values at which Fis, GyrAB, GyrI, and TopA control the DNA supercoiling level and at which Fis stimulates the expression of the stable RNAs. This yields a total of 1296 different PL models, for which we first test the ability to capture the steady states corresponding to stationary and exponential phase. All the PL models have two stable steady states, one corresponding to stationary-phase conditions, with the carbon starvation signal present and a predicted low level of stable RNAs, indicating the absence of cellular growth. In the other steady state the signal is switched off, but in many cases the state is not representative for exponential-phase conditions since a low level of stable RNAs is predicted. Because we are interested only in biologically meaningful models, we have studied which sets of inequality constraints confer a good prediction of the exponential-phase conditions. It appears that one simple inequality is sufficient to discriminate between the models. In particular, the threshold value at which Fis protein inhibits the expression of *gyrAB* should be higher than the concentration needed to activate the production of stable RNAs, to allow the accumulation of the latter in presence of glucose. This relative ordering of Fis binding affinities for the promoter regions of the stable RNAs and the gyrase has never been described in the experimental literature and thus constitutes an interesting prediction for further investigation.

The addition of the inequality constraint between the above-mentioned thresholds for Fis reduces the admissible sets of inequality constraints to 258. For all 258 PL models, we have run a qualitative simulation of the response to carbon starvation in exponential phase. All the models tested predict that the accumulation of RpoS to high levels is essential for the downregulation of the DNA supercoiling level, as tested by means of model checking. RpoS may thus be responsible for the relaxation of DNA topology at the entry into stationary phase. On the contrary, in none of the 258 models, the accumulation of RpoS to high levels or a low DNA supercoiling level are necessary for the cell's growth arrest. Only the accumulation of the complex cAMP-CRP is essential for this process, confirming the key role played by the mutual inhibition of *fis* and *crp*.

These results thus allow us to clarify our picture of the carbon starvation response by assigning a role to the individual global regulator in the network. A striking result is that the 258 PL models predict essentially the same effect of RpoS accumulation on the DNA supercoiling level and the downregulation of the stable RNAs. This suggests that these properties are robust for a large range of parameter values controlling the DNA supercoiling level.

## 6.7 Conclusions

The modeling of the regulatory networks controlling the response of bacteria to external perturbations leads to large and complex systems of nonlinear ODEs. These models are difficult to study in the absence of quantitative information on parameter

values, which has motivated the use of various approximations adapted to typical response functions and timescale hierarchies found in genetic and metabolic regulation. In this chapter, we have reviewed one such model reduction approach, based on QSS and PL approximations. Under the condition that the dominant timescale of interest is that of protein synthesis and degradation, which involves sigmoidal response functions that can be suitably approximated by step functions; these approximations are expected to give rise to simplified models of complex networks with dynamics close to that of classical ODE models (Ropers et al. (2011), see also Chaves et al. (2006), Davidich and Bornholdt (2008)).

The PL models obtained after model reduction have been well-studied in the literature, following their original introduction by Glass and Kauffman (Batt et al. 2007; Edwards 2000; Ghosh and Tomlin 2004; Glass and Kauffman 1973; Gouzé and Sari 2002; Mestl et al. 1995; Plahte and Kjöglum 2005). The qualitative dynamics of the PL models can be represented by a state-transition graph, consisting of states and transitions between states. A major result is that this graph is invariant for large sets of parameter values, defined by inequality constraints that can be inferred from the experimental literature. Moreover, the state transition graph can be computed from the inequality constraints by means of simple, symbolic rules. In order to support its application to large and complex genetic regulatory networks, the analysis of the PL models has been implemented in the computer tool GNA. As the graphs become too large to be analyzed by hand, GNA can be coupled with model-checking tools for the automatic verification of dynamic properties of the network. This approach can be particularly useful for the analysis of incompletely specified models, as summarized in Fig. 6.4.

The use of PL models is justified by the intuition that, to a first approximation, genes can be considered logical switches that transform continuous inputs – i.e., the concentration of regulatory proteins – into discrete outputs – i.e., the activation state of the genes (Sugita 1963; Yuh et al. 1998). Instead of developing this intuition for models with continuous time and concentration variables, one could also decide to employ discrete models. The major example of this approach is the *Boolean network* formalism developed by Kauffman, Thomas, and others (Kauffman 1969, 1993; Thomas 1973; Thomas and d’Ari 1990). The application of Boolean networks rests on the assumption that a gene is either active or inactive, and that genes change their activation state synchronously. For the purpose of modeling actual genetic regulatory networks, these assumptions are usually too strong. In response to this problem, more general formalisms with multivalued activation states and asynchronous transitions have been proposed and successfully applied to the analysis of complex developmental regulatory networks (Gonzalez et al. 2006; Mendoza et al. 1999; Sánchez and Thieffry 2003). The advantage of Boolean networks and their generalizations is that they provide a convenient way to express the logic of gene expression regulation. However, they have difficulty in treating dynamic properties of genetic regulatory networks taking place at the threshold of activation or inactivation of a gene, where steady states may be located (Casey et al. 2006).

PL models may not be appropriate when the systems under study do not involve sigmoidal response functions, and are therefore not well approximated by

step functions. This is particularly so for metabolism and cell signaling, where Michaelis-Menten and mass-action kinetics are common. Other types of approximations may apply in these cases though, such as lin-log models (Heijnen 2005), power-law models (Savageau 2001), and piecewise multi-affine models (Belta and Habetts 2006). Even when step-functions approximations are appropriate, they may not be sufficiently precise. The methods described in this chapter strive at capturing the qualitative dynamics of the networks, but when quantitative precision is sought, it may be necessary to use less drastic approximations of the sigmoidal response functions.

In this chapter, we have reviewed the use of PL models in the context of the modeling of the network of global regulators involved in the carbon starvation response in *E. coli*. Qualitative simulation has been used to obtain predictions of the behavior of networks which is currently not yet well understood by biologists. While some of the predictions help clarifying the role of particular regulatory mechanisms (the mutual inhibition of *fis* and *crp*), others concern phenomena that have not yet been experimentally investigated (the occurrence of damped oscillations after a nutrient upshift and the relative ordering of Fis binding affinities for different promoter regions). The latter two predictions are particularly interesting from a biological point of view, because they generate new questions and suggest further experiments.

The basic motivation for the use of the PL models is the current absence of precise and quantitative information on kinetic parameters. The advantage of the reduction of ODE models to PL models is that they allow a quick scan of the qualitative dynamics of the system, without numerical information on parameter values. This provides a first insight into the global dynamics of the system, which is interesting in itself but also yields structural and parametric constraints that may guide the development of more detailed kinetic models. In particular, it may orient the difficult problems of system identification and parameter estimation from experimental data (Bettenbrock et al. 2006; Gardner et al. 2003; Kuepfer et al. 2007; Quach et al. 2007; Moles et al. 2003; Porreca et al. 2008; Ronen et al. 2002; van Riel and Sontag 2006; Vilela et al. 2008; Westra et al. 2007; Zwolak et al. 2005).

**Acknowledgment** VB, DR, JG, HdJ were supported by the European commission under project EC-MOAN (FP6-2005-NEST-PATH-COM/043235). PM was partially supported by FCT program (PhD grant SFRH/BD/32965/2006 to PTM) and PDTC program (project PTDC/EIA/71587/2006).

## References

- T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, 181 (20): 6361–6370, 1999.
- M. Antonioti, A. Policriti, N. Ugel, and B. Mishra. Model building and model checking for biochemical processes. *Cell Biochem. Biophys.*, 38 (3): 271–286, 2003.
- C. A. Ball, R. Osuna, K. C. Ferguson, and R. C. Johnson. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J. Bacteriol.*, 174 (24): 8043–8056, 1992.

- G. Batt, D. Ropers, H. de Jong, J. Geiselman, R. Mateescu, M. Page, and D. Schneider. Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics*, 21 (Suppl 1): i19–i28, 2005.
- G. Batt, B. Yordanov, R. Weiss, and C. Belta. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics*, 23 (18): 2415–2422, 2007.
- G. Batt, H. de Jong, M. Page, and J. Geiselman. Symbolic reachability analysis of genetic regulatory networks using discrete abstractions. *Automatica*, 44 (4): 982–989, 2008.
- C. Belta and L. C. G. J. M. Habets. Controlling a class of nonlinear systems on rectangles. *IEEE Trans. Autom. Control*, 51 (11): 1749–1759, 2006.
- G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas’ asynchronous logical approach with temporal logic. *J. Theor. Biol.*, 229 (3): 339–348, 2004.
- K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, and E. D. Gilles. A quantitative approach to catabolite repression in *Escherichia coli*. *J. Biol. Chem.*, 281 (5): 2578–2584, 2005.
- K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, and E. D. Gilles. A quantitative approach to catabolite repression in *Escherichia coli*. *J. Biol. Chem.*, 281 (5): 2578–2584, 2006.
- M. Calder, V. Vyshemirsky, D. Gilbert, and R. Orton. Analysis of signalling pathways using the PRISM model checker. In G. Plotkin, editor, *Proc. of CMSB*, pages 179–190, Edinburgh, Scotland, 2005.
- R. Casey, H. de Jong, and J.-L. Gouzé. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. *J. Math. Biol.*, 52 (1): 27–56, 2006.
- N. Chabrier-Rivier, M. Chiaverini, V. Danos, F. Fages, and V. Schächter. Modeling and querying biomolecular interaction networks. *Theor. Comput. Sci.*, 325 (1): 25–44, 2004.
- C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.*, 79 (1): 53–73, 2002.
- M. Chaves, E. D. Sontag, and R. Albert. Methods of robustness analysis for Boolean models of gene control networks. *IET Syst. Biol.*, 153 (4): 154–167, 2006.
- K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell*, 15 (8): 3841–3862, 2004.
- A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV2: An OpenSource tool for symbolic model checking. In D. Brinksma and K. G. Larsen, editors, *14th International Conference on Computer Aided Verification (CAV 2002)*, volume 2404 of *LNCS*, pages 359–364. Springer, Berlin, 2002.
- E. M. Clarke, O. Grumberg, and D. A. Peled. *Model Checking*. MIT, Boston, MA, 1999.
- A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, London, revised edition, 1995.
- M. Davidich and S. Bornholdt. The transition from differential equations to Boolean networks: A case study in simplifying a regulatory network model. *J. Theor. Biol.*, 255: 269–277, 2008.
- H. de Jong and M. Page. Search for steady states of piecewise-linear differential equation models of genetic regulatory networks. *ACM/IEEE Trans. Comput. Biol. Bioinform.*, 5 (2): 208–222, 2008.
- H. de Jong and D. Ropers. Strategies for dealing with incomplete information in the modeling of molecular interaction networks. *Brief. Bioinform.*, 7 (4): 354–363, 2006.
- H. de Jong, J. Geiselman, C. Hernandez, and M. Page. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19: 336–344, 2003.
- H. de Jong, J. Geiselman, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis*. *B. Math. Biol.*, 66 (2): 261–299, 2004a.
- H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselman. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *B. Math. Biol.*, 66 (2): 301–340, 2004b.
- R. Edwards. Analysis of continuous-time switching networks. *Phys. D*, 146 (1–4): 165–199, 2000.
- J. Fisher, N. Piterman, A. Hajnal, and T. A. Henzinger. Predictive modeling of signaling crosstalk during *C. elegans* vulval development. *PLoS Comput. Biol.*, 3 (5): e92, 2007.

- T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403 (6767): 339–342, 2000.
- T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301 (5629): 102–105, 2003.
- R. Ghosh and C. J. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-Notch protein signalling. *IET Syst. Biol.*, 1 (1): 170–183, 2004.
- L. Glass and S. A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.*, 39 (1): 103–129, 1973.
- A. Gonzalez Gonzalez, A. Naldi, L. Sánchez, D. Thieffry, and C. Chaouiya. GINsim: a software suit for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, 84 (2): 91–100, 2006.
- J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dyn. Syst.*, 17 (4): 299–316, 2002.
- R. M. Gutierrez-Ríos, J. A. Freyre-Gonzalez, O. Resendis, J. Collado-Vides, M. Saier, and G. Gosset. Identification of regulatory network topological units coordinating the genome-wide transcriptional response to glucose in *Escherichia coli*. *BMC Microbiol.*, 7: 53–53, 2007.
- A. Halász, V. Kumar, M. Imielinski, C. Belta, O. Sokolsky, S. Pathak, and H. Rubin. Analysis of lactose metabolism in *E. coli* using reachability analysis of hybrid systems. *IET Syst. Biol.*, 1 (2): 130–48, 2007.
- T. Hardiman, K. Lemuth, M. A. Keller, M. Reuss, and M. Siemann-Herzberg. Topology of the global regulatory network of carbon limitation in *Escherichia coli*. *J. Biotechnol.*, 132: 359–374, 2007.
- J. J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.*, 91 (5): 534–545, 2005.
- R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman & Hall, New York, 1996.
- G. W. Huisman, M. M. Siegel, D. A., Zambrano, and Kolter R. Morphological and physiological changes during stationary phase. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, editors, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pages 1672–1682. ASM, Washington D.C., 1996.
- H. Ishizuka, A. Hanamura, T. Inada, and H. Aiba. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. *EMBO J.*, 13 (13): 3077–3082, 1994.
- S. A. Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, New York, 1993.
- S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22 (3): 437–467, 1969.
- E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.*, 23 (8): 975–982, 2005.
- K. W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11 (1): 84–97, 2001.
- A. Kremling, S. Kremling, and K. Bettenbrock. Catabolite repression in *Escherichia coli* - a comparison of modelling approaches. *FEBS J.*, 276: 594–602, 2009.
- L. Kuepfer, M. Peter, U. Sauer, and J. Stelling. Ensemble modeling for analysis of cell signaling dynamics. *Nat. Biotechnol.*, 25 (9): 1001–1006, 2007.
- J.-C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *Proc. Nat. Acad. Sci. USA*, 100 (12): 7051–7056, 2003.
- L. Mendoza, D. Thieffry, and E. R. Alvarez-Buylla. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis. *Bioinformatics*, 15 (7–8): 593–606, 1999.
- T. Mestl, E. Plahte, and S. W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.*, 176 (2): 291–300, 1995.
- T. Millat, E. Bullinger, J. Rohwer, and O. Wolkenhauer. Approximations and their consequences for dynamic modelling of signal transduction pathways. *Math. Biosci.*, 207 (1): 40–57, 2007.

- C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.*, 13 (11): 2467–2474, 2003.
- P. T. Monteiro, D. Ropers, R. Mateescu, A. T. Freitas, and H. de Jong. Temporal logic patterns for querying dynamic models of cellular interaction networks. *Bioinformatics*, 24: i227–i233, 2008.
- M. S. Okino and M. L. Mavrouniotis. Simplification of mathematical models of chemical reaction systems. *Chem. Rev.*, 98 (2): 391–408, 1998.
- J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol.*, 22 (8): 400–405, 2004.
- E. Pecou. Splitting the dynamics of large biochemical interaction networks. *J. Theor. Biol.*, 232 (3): 375–384, 2005.
- E. Plahte and S. Kj glum. Analysis and generic properties of gene regulatory networks with graded response functions. *Phys. D*, 201 (1): 150–176, 2005.
- R. Porreca, S. Drulhe, H. de Jong, and G. Ferrari-Trecate. Structural identification of piecewise-linear models of genetic regulatory networks. *J. Comput. Biol.*, 15 (10): 1365–1380, 2008.
- T. S. Pratt, T. Steiner, L. S. Feldman, K. A. Walker, and R. Osuna. Deletion analysis of the *fis* promoter region in *Escherichia coli*: antagonistic effects of integration host factor and Fis. *J. Bacteriol.*, 179 (20): 6367–6377, 1997.
- M. Quach, N. Brunel, and F. d’Alch  Buc. Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference. *Bioinformatics*, 23 (23): 3209–3216, 2007.
- Y. Reyes-Dominguez, G. Contreras-Ferrat, J. Ramirez-Santos, J. Membrillo-Hernandez, and M. C. Gomez-Eichelmann. Plasmid DNA supercoiling and gyrase activity in *Escherichia coli* wild-type and *rpoS* stationary-phase cells. *J. Bacteriol.*, 185 (3): 1097–1100, 2003.
- J. M. Rohwer, N. D. Meadow, S. Roseman, H. V. Westerhoff, and P. W. Postma. Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements *in vitro*. *J. Biol. Chem.*, 275 (45): 34909–34921, 2000.
- M. Ronen, R. Rosenberg, B. I. Shraiman, and U. Alon. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci. USA*, 99 (16): 10555–10560, 2002.
- D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems*, 84 (2): 124–152, 2006.
- D. Ropers, V. Baldazzi, and H. de Jong. Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli*. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8 (1): 166–181, 2011.
- M. R. Roussel and S. J. Fraser. Invariant manifold methods for metabolic model reduction. *Chaos*, 11 (1): 196–206, 2001.
- L. S nchez and D. Thieffry. Segmenting the fly embryo: A logical analysis of the pair-rule cross-regulatory module. *J. Theor. Biol.*, 224 (4): 517–537, 2003.
- M. A. Savageau. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos*, 11 (1): 142–159, 2001.
- B. Schoeberl, C. Eichler-Jonsson, E.-D. Gilles, and G. M ller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, 20 (4): 370–375, 2002.
- J.-A. Sepulchre, S. Reverchon, and W. Nasser. Modeling the onset of virulence in a pectinolytic bacterium. *J. Theor. Biol.*, 44 (2): 239–257, 2007.
- M. Sugita. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent: II. The idea of a molecular automaton. *J. Theor. Biol.*, 4: 179–192, 1963.
- R. Thomas. Boolean formalization of genetic control circuits. *J. Theor. Biol.*, 42 (3): 563–585, 1973.
- R. Thomas and R. d’Ari. *Biological feedback*. CRC, Boca Raton, FL, 1990.
- R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory: II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11 (1): 180–195, 2001.

- A. Usseglio Viretta and M. Fussenegger. Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*. *Biotechnol. Prog.*, 20 (3): 670–678, 2004.
- N. A. W. van Riel and E. D. Sontag. Parameter estimation in models combining signal transduction and metabolic pathways: The dependent input approach. *IET Syst. Biol.*, 153 (4): 263–274, 2006.
- M. Vilela, I. Chou, S. Vinga, A. Vasconcelos, E. Voit, and J. Almeida. Parameter optimization in S-system models. *BMC Syst. Biol.*, 2: 35, 2008.
- R. L. Westra, G. Hollanders, G. J. Bex, M. Gyssens, and K. Tuyls. The identification of dynamic gene-protein networks. In K. Tuyls, R. Westra, Y. Saeys, and A. Nowé, editors, *Proc. KDEC2006*, volume 4366 of *LNCS*, pages 157–170. Springer, Berlin, 2007.
- C.-H. Yuh, H. Bolouri, and E. H. Davidson. Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science*, 279: 1896–1902, 1998.
- J. W. Zwolak, J. J. Tyson, and L. T. Watson. Parameter estimation for a mathematical model of the cell cycle in frog eggs. *J. Comput. Biol.*, 12 (1): 48–63, 2005.

# Chapter 7

## Modeling Antibiotic Resistance in Bacterial Colonies Using Agent-Based Approach

James T. Murphy and Ray Walshe

### 7.1 Introduction

Multi-drug resistance in *Staphylococcus aureus* bacteria has become a major health care challenge in recent decades. Infections with *S. aureus* had a mortality rate of over 80% before the introduction of antibiotics in the early 1940s (Skinner and Keefer 1941). The first  $\beta$ -lactam antibiotic, penicillin, was introduced into clinical use during the early 1940s and dramatically reduced the mortality rate associated with these infections (Chain et al. 1993). However, the widespread use of  $\beta$ -lactam antibiotics led to the rapid expansion of resistant strains of bacteria. As a result, today greater than 95% of all *S. aureus* isolates have been found to possess resistance to penicillin, and methicillin resistance is estimated to be in 40–60% of clinical isolates in the USA and the UK (Levy and Marshall 2004; Neu 1992).

Advances in cell and molecular biology techniques in the latter half of the twentieth century have led to rapid increases in our knowledge about the basic cellular processes involved in antibiotic resistance. This development has allowed a more fine-grained approach to be taken in investigating the spread of resistance in populations of bacteria. However, the overall population response to antibiotic treatment is often a function of a diverse range of interacting components. A sound theoretical understanding of the systems of interactions taking place (e.g. between antibiotic molecules and cell surface proteins) is required to develop strategies to minimise the spread of antibiotic resistance. The rapid development in the areas of pharmacokinetics and pharmacodynamics studies in recent years is a response to this need to understand the complex dynamics that contribute to the bacterial response to drug treatment (Ambrose et al. 2007).

There has also been a strong development of studies into the population and growth dynamics of bacterial populations using various computational modeling approaches. The most common of these are the mathematical models which describe

---

J.T. Murphy (✉)  
Centre for Scientific Computing and Complex Systems Modeling, School of Computing,  
Dublin City University, Dublin 9, Ireland  
e-mail: jamurphy@computing.dcu.ie

the population as a whole using state variables. These have been important for developing insights into parameters at the population level that influence the development of the colony (Lacasta et al. 1999). They allow an integrated view of colony development to be realised, and to identify key determinants of population growth and development.

However, models that use global parameters for a population imply that they are in a homogeneous, mixed environment. In nature, bacteria often form highly heterogeneous colonies where there can be significant localised variations in the chemical environment such as nutrient availability, temperature, ion concentrations and pH (Devine 2004). A different approach, which allows the heterogeneity of populations to be explicitly modeled, is the agent-based (or individual-based) modeling approach. For this, the individual bacterial cells represent the fundamental units of the simulation. This approach where the parameters are defined for the bacterial cells rather than for the population as a whole is called a “bottom-up” approach. The properties of a colony emerge from the set of interactions of a population of heterogeneous bacterial agents.

This approach is what was taken in our research to model the interactions of bacterial cells with antibiotic molecules. An agent-based model was developed, called the Micro-Gen Bacteria Simulator, which simulates the life cycle of bacteria growing in culture and their interactions with various molecules including antibiotics (Walshe 2006; Murphy and Walshe 2006; Murphy et al. 2007). The model engine was designed to be adaptable to represent different species and strains of bacteria using basic cellular information. For our research, a model of methicillin-resistant *Staphylococcus aureus* (MRSA) bacteria was developed, since they are currently of great clinical significance and there is a wealth of biological information available about them.

The model incorporates representations of the two main antibiotic resistance mechanisms characteristic of MRSA ( $\beta$ -lactamase enzymes and PBP2a, see below). It includes a quantitative model of the interactions between  $\beta$ -lactam antibiotics and MRSA bacteria using kinetic rules for these reactions derived from experimental studies (Zygmunt et al. 1992; Lu et al. 1999). By simulating the individual molecular interactions of antibiotic molecules and bacteria, and scaling this up to large population sizes using the agent-based approach, Micro-Gen can be used to explore the emergent dynamics that contribute to antibiotic resistance in bacterial populations.

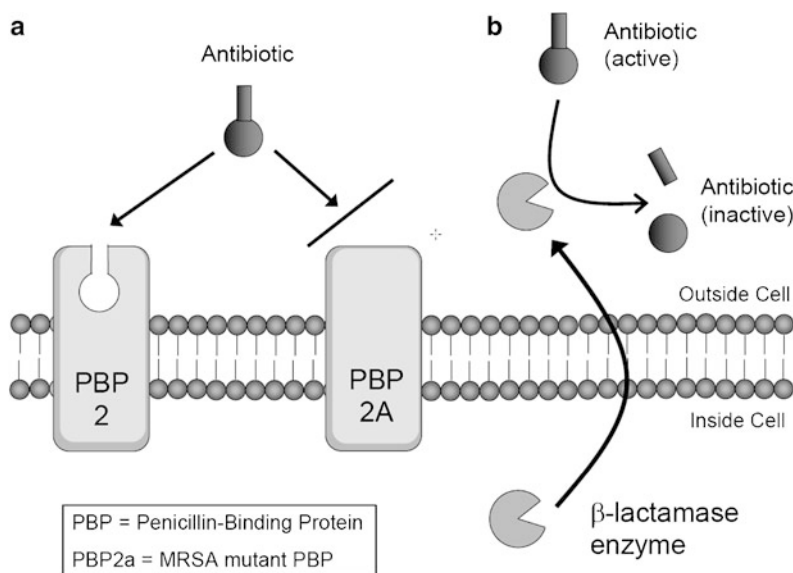
### **7.1.1 MRSA Antibiotic Resistance Mechanisms**

Antibiotic resistance refers to the ability of a microorganism to resist the effects of an antibiotic. Examples of mechanisms of antibiotic resistance include the synthesis of antibiotic-degrading enzymes (e.g.  $\beta$ -lactamase) and modifications to drug targets such as the penicillin-binding proteins (PBPs) in bacterial cell membranes. There is a complex range of interacting factors that influence the development and

spread of antibiotic resistance in pathogenic bacteria. Even though the appearance of resistant strains of bacteria is thought to be an ancient evolutionary event, the fitness cost associated with resistance mechanisms limited their proliferation before the introduction of antibiotics (Wright 2007). The widespread use of antibiotics has resulted in a significant positive selective pressure for resistant strains, particularly in the clinical setting (Levy and Marshall 2004).

There are several resistance mechanisms that have been observed in bacteria in response to antibiotic exposure. These include mechanisms to limit the uptake of antibiotic molecules into the bacterial cells. For example, the deletion of porin proteins in gram-negative bacteria to block the passage of antibiotic molecules through the bacterium's outer membrane, and the activation of efflux exporter proteins to "pump out" the antibiotic (Walsh 2000; Fisher et al. 2005). Perhaps the most clinically significant resistance mechanisms, with respect to  $\beta$ -lactam antibiotics, in gram-positive pathogens such as *S. aureus* bacteria are the expression of enzymes called  $\beta$ -lactamases, and alterations to the molecular targets (PBPs) of the antibiotics (Fig. 7.1).

Penicillin resistance in *S. aureus* is mediated by the production and release of an enzyme, called  $\beta$ -lactamase, which hydrolytically cleaves the  $\beta$ -lactam ring structure present in penicillin and other  $\beta$ -lactam antibiotics.  $\beta$ -lactamase was first discovered in *Escherichia coli* bacteria in 1940, and  $\beta$ -lactamase-expressing *S. aureus* bacteria were also isolated soon afterwards (Abraham and Chain 1940; Bondi



**Fig. 7.1** Diagram of the two principal antibiotic resistance mechanisms observed in MRSA bacteria. (a) Expression of alternate form of penicillin-binding protein PBP2, called PBP2a, with reduced binding affinity for antibiotic. (b) Production and release of  $\beta$ -lactamase enzyme which cleaves and inactivates antibiotic molecules

and Dietz 1945; Kirby 1944).  $\beta$ -lactamases are thought to have evolved long before the clinical introduction of  $\beta$ -lactam antibiotics. However, they only became broadly distributed across many bacterial species under selective pressure from widespread antibiotic use (Fisher et al. 2005).

The interaction between  $\beta$ -lactams in the environment and a cell surface signal-transducer protein, BlaR1, triggers the process whereby expression of  $\beta$ -lactamase occurs (Fuda et al. 2005; Lewis et al. 1999). Most of the  $\beta$ -lactamase enzyme produced by the cell is secreted into the extracellular milieu, while some remains bound to the cytoplasmic membrane of the cell (Nielsen and Lampen 1982). Re-repression of  $\beta$ -lactamase expression occurs when the antibiotic concentration in the environment decreases and BlaR1 is no longer auto-activated (Zhang et al. 2001).

MRSA strains were first discovered soon after the introduction of methicillin in 1959 (Fuda et al. 2005). Methicillin had been introduced to treat infections of penicillin-resistant *S. aureus* which had become a significant health concern at that stage. MRSA strains were isolated as early as 1961, and they have steadily increased in frequency since then in response to selective pressure (Eriksen 1961). MRSA bacteria contain a gene called *mecA*, which encodes a penicillin-binding protein, PBP2a, which circumvents the mode of action of  $\beta$ -lactam antibiotics. Normal *S. aureus* cells produce four types of membrane-bound transpeptidase proteins called penicillin-binding proteins (PBPs 1–4), which assemble and regulate the final stages of cell wall biosynthesis. The mode of action of the  $\beta$ -lactam antibiotics involves the acylation of a catalytic residue in the transpeptidase active site of PBPs which results in the inhibition of their cell-wall cross-linking function, a crucial step in cell wall assembly during bacterial cell division.

Therefore, when the bacterial cell is unable to correctly assemble the cell wall due to binding of the PBPs by antibiotic, this can result in inhibition of cell division (bacteriostatic effect) or even cell death (bactericidal effect). However, the product of the *mecA* gene, PBP2a, does not bind the  $\beta$ -lactam moiety readily because the approach to the active site is sterically encumbered. When an MRSA organism is exposed to  $\beta$ -lactam antibiotics, PBP2a confers resistance by supplementing its transpeptidase activity (cell-wall cross-linking) to the transglycosylase function of native PBPs during cell wall synthesis (Fuda et al. 2005). As a result, cell wall assembly can continue to occur even in the presence of antibiotic.

### 7.1.2 Overview of Modeling Approaches

The power of computational modeling approaches lies not so much in their ability to make predictions (since some degree of experimental validation is needed to confirm any predictions) but rather in their ability to give new insights into the underlying mechanistic basis for the observed biological phenomena. A computational model was defined by Volker Grimm as a “purposeful representation” of an entity or system whose “purpose is to capture the essence of a problem and explore different solutions of it” (Grimm 1999). The most important role of a model is therefore

to aid in our understanding of a particular process. From this perspective, all the different modeling approaches share the same principal aim, though they may differ in the assumptions and tools that are used.

The most common mathematical approaches to modeling bacterial population growth have been the use of *ordinary differential equations (ODEs)* and partial differential equations (PDEs). ODE models are often used in systems biology because they are computationally efficient and mathematically robust, and can be used to develop an integrated view of biological systems. Extensions to the basic ODE methods have also been developed over the years such as stochastic ODEs and compartmentalised ODE models. PDEs are used to model processes that have spatial as well temporal dependencies.

Mathematical population models use global parameters or state variables to describe the growth and development of a bacterial colony as a unit (Grimson and Barker 1994; Lacasta et al. 1999). These “top-down” approaches have the advantage that they are computationally efficient and less parameter-rich than more low-level approaches. However, with state variable approaches it is sometimes difficult to trace back the system behaviour to the behaviour of the individual agents. For example, this approach does not explicitly explain the underlying factors that lead to the population exhibiting a particular growth rate or carrying capacity (Grimm 1999). However, it is important because it provides an appropriate integrated view of the population behaviour.

Another modeling technique commonly used in systems biology of bacterial colonies is the theory of *cellular automata (CA)* (Ben-Jacob et al. 1994). This approach is a powerful tool for representing both temporal and spatiotemporal processes in biological systems. In CA models, the environment of the model is represented by a discrete lattice/grid where the states of the components evolve synchronously in discrete time steps according to a set of rules. The CA simulation Conways Game of Life was one of the first examples of computer applications in biology (Gardner 1970). This model consisted of randomly placed cells on a square lattice and simulated birth, death and interactions according to pairwise interaction rules which used Boolean logic conditions. CA methods have continued to be developed since then and applied to a diverse range of problems in computational systems biology (Materi and Wishart 2007).

### 7.1.3 Agent-Based Modeling Approach

An alternative approach to modeling bacterial growth and development, and the main subject of this chapter, is the agent-based (or individual-based) modeling approach (Ginovart et al. 2002; Kreft et al. 1998). The distinguishing characteristic of the agent-based approach is that the properties of the individual cells, rather than the colony as a whole or a subgroup of it, are modeled. This “bottom-up” approach allows a finer-grained analysis than the other techniques, connecting local changes at the cellular level to the overall patterns of population growth. The agent-based

approach shares some of the strengths of the cellular automata modeling approach, in which it is able to explicitly model both temporal and spatiotemporal processes. For this reason, it is particularly amenable to modeling processes such as chemotaxis, diffusion and pattern formation in bacterial colonies. However, it represents an even finer-grained approach than CA in which the individual biological entities being modeled are explicitly represented by unique software agents.

Complex agent-based systems consist of many similar and simple components. Often the system as a whole has a complex behaviour that is more than the sum of its constituent parts. Agents can be used to conceptualise and implement such a software modeling application. An agent-oriented approach lends itself naturally to the complexity of modeling bacterial interactions. For the purposes of this simulation, the definition of an agent will be adopted as given by Jennings et al. (1998) in which an agent is described as a “computer based system, situated in some environment which is capable of flexible, autonomous action to meet its goals”. Within this definition, there are three fundamental concepts – situatedness, autonomy and flexibility.

“Situatedness” in this context means that the agent is in some environment from which it receives sensory input and is capable of modifying the environment. So, for example, in terms of a bacterial simulation, a bacterium in an agar plate senses nutrient and can consume it. “Autonomy” in this context implies that the system should be able to act independently from human intervention. The system has local control over its own state, i.e. once it is set up and running no intervention is required. “Flexibility” means that the system should be re-active, pro-active and social. *Re-active*: the system should respond in a timely fashion to changes within the environment. *Pro-active*: the system should bring about changes in the environment to meet its design objectives. *Social*: the system should interact, collaborate or compete, where necessary to complete its design objectives.

The theory of autonomous agents is a useful approach for the modeling of bacterial cell colonies as it allows large-scale population models to be derived from simple rules dictating the growth and interactions of the individual bacterial cells of the population. The model produces global information about population growth in different environmental conditions using basic information about the cell biology of bacteria. The agent-based approach was chosen to explicitly model the heterogeneity in environmental conditions, for example between the interior and exterior of a colony, and between individual bacterial cells. This would not be so amenable using another approach such as a simpler mass action model. In complex microbial communities, there can be highly heterogeneous localised niches where the chemistry varies significantly over small distances, and the agent-based approach allows us to take this into account. It could also be used to model other mechanisms such as phase and antigenic variation in bacteria that result in heterogenic phenotypes within a clonal population (van der Woude and Baumler 2004).

Some of the challenges of this approach include the fact that it can sometimes require more parameters than a state variable approach since the individual entities are explicitly modeled and it also may become susceptible to empirical knowledge (Ambrose et al. 2007). In other words, it is sometimes difficult to obtain accurate

parameters for individual cells. However, advances in experimental techniques related to studying individual cells will help in building more accurate models at the cellular level (Elfving et al. 2004; Niven et al. 2006). Using appropriate aggregation of parameters, and cognisant of its limitations, the agent-based approach can be used as a powerful tool for tracing back system behaviour to that of its individual components.

The agent-based approach and higher-level mathematical approaches are not mutually exclusive but rather it is envisaged that they should complement each other in studies of population dynamics. Mathematical models allow theories at the systems level to be developed by providing a general conceptual framework for the population as a whole. Meanwhile, the agent-based approach allows the overall system properties to be related back to the individuals of the population. Agent-based models can be more computationally intensive than high-level mathematical approaches such as ODEs, because each individual of a population is explicitly modeled. However, in cases where the population expresses a high degree of heterogeneity, both spatially and between individuals, the agent-based approach can represent a powerful tool for exploring how this heterogeneity contributes to the overall system dynamics.

## 7.2 Micro-Gen Bacterial Simulator

In this section, there is a detailed overview of the main structure and components of the agent-based software model *Micro-Gen*, developed to model the growth and interactions of bacterial cells with antibiotics in vitro (Murphy and Walshe 2006; Murphy et al. 2007). It was developed from existing work in the laboratory carried out previously on an agent-based model called the Bacteria–Antibiotic Interaction Tool (BAIT) (Walshe 2006). BAIT consisted of a simple model of bacterial growth and interactions with antibiotic molecules in a discrete two-dimensional grid environment using the Java programming language. It demonstrated the feasibility of this approach for examining the individual dynamics of antibiotic molecules interacting with bacterial cells in culture. *Micro-Gen* represents a significant expansion and re-design of the original BAIT tool, to build a more realistic representation of bacterial growth and development in culture and a quantitative model of their interactions with antibiotics.

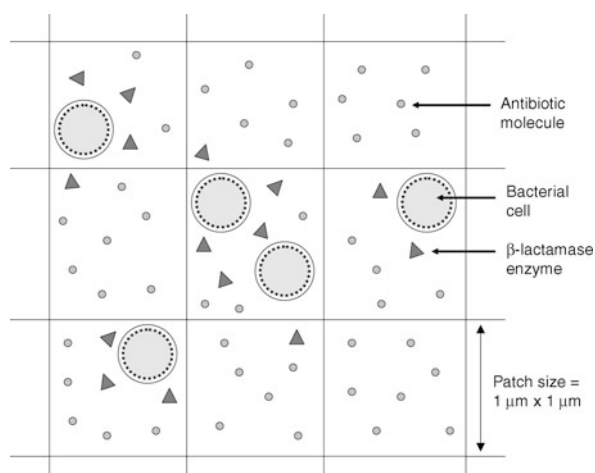
*Micro-Gen* is coded in the C++ object-oriented programming language. The individual microorganisms are represented by software agents which store physical traits of the bacterial cells as well as behavioural rules associated with them. The modular nature of the program means that functionalities/characteristics specific to particular bacterial species can be readily incorporated into the basic cellular model. A key component of the model is the ability to quantitatively model antibiotic molecules and their interactions with the bacterial cells. These interactions between the antibiotic molecules and targets in the bacterial cell are governed by defined kinetic parameters specific to the type of antibiotic and bacterial strain being

modeled. This allows a quantitative model of antibiotic interactions with bacteria to be built up and their pharmacokinetic properties to be investigated.

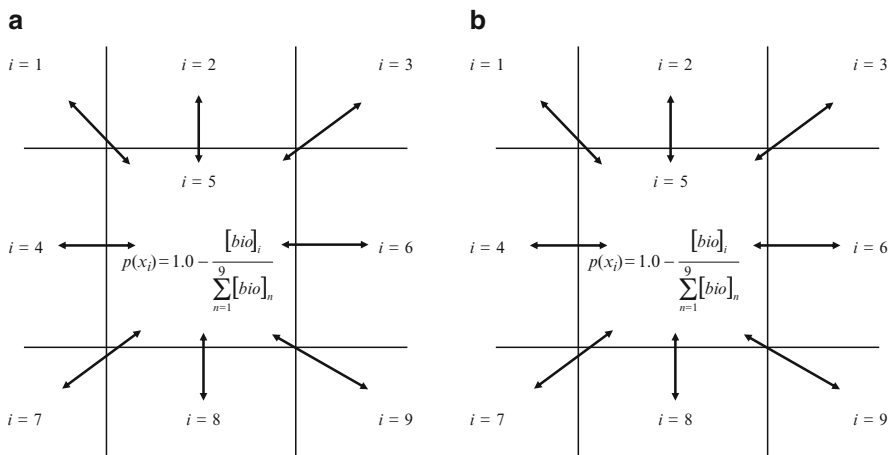
The model incorporates two important antibiotic resistance mechanisms used by bacterial cells against antimicrobial agents which form the cornerstone of the antibiotic arsenal: special enzymes released by bacteria, called  $\beta$ -lactamases, which degrade the antibiotic molecules, and reduced binding affinities between the antibiotics and receptors in the bacterial cells (see Sect. 7.1.1). These antibiotic resistance mechanisms are of great clinical concern as their development and spread across many species of bacteria has led to the erosion of the efficacy of many commonly prescribed antibiotics, in particular penicillin and its derivatives.

### 7.2.1 Environment

The simulated culture environment consists of a discrete, two-dimensional grid containing diffusible elements such as antibiotics, nutrients (glucose) and  $\beta$ -lactamase enzymes (Fig. 7.2). The environment was restricted to a two-dimensional plane to minimise the computational burden of the program. Each discrete grid position is referred to as a “patch” (to differentiate it from a bacterial “cell”) and contains variables to record the levels of the various molecular components in it. It also includes pointers to bacterial agents that currently occupy the patch. A discretised implementation of Fick’s first law of diffusion is used to calculate the movement of molecules between adjacent patches down local concentration gradients (Ginovart et al. 2002). The amount of substance exchanged between two adjacent patches



**Fig. 7.2** Overview of two-dimensional grid environment in *Micro-Gen*. Each grid element is referred to as a “patch” and contains various simulation components including bacterial cells,  $\beta$ -lactamase enzymes, antibiotic molecules and nutrients



**Fig. 7.3** Diffusion and overcrowding algorithms. **(a)** Diffusion algorithm (Fick's First Law of diffusion) applied to molecules.  $D$  diffusion coefficient,  $\Delta Mol$  concentration difference. **(b)** Overcrowding algorithm applied to bacterial cells.  $p(x_i)$  probability of bacterial agent moving to patch  $i$ ,  $[bio]_i$  bacterial biomass in patch  $i$

is calculated as the concentration difference ( $\Delta Mol$ ) multiplied by a user-defined diffusion coefficient,  $D$  (Fig. 7.3a). When patches are diagonally adjacent to one another, a diffusion rate modifier ( $1/\sqrt{2}$ ) is applied.

## 7.2.2 Bacterial Agents

The bacterial agents are autonomous entities with a set of behavioural rules that determine how they interact with the environment and parameters associated with them recording details such as their energy state (or nutrient reserve). In order to keep the model simple, the internal subcellular processes of the bacteria are not explicitly modeled. This minimises the number of parameters associated with each agent, thus optimising the performance of the program and avoiding the problems associated with a parameter-rich model (Ambrose et al. 2007). The main input parameters for the bacterial agents, with some sample values for *S. aureus*, are listed in Table 7.1.

The bacterial cells are assumed to be simple, independent entities that passively absorb nutrients from the environment and grow and divide asexually as they accumulate more nutrients. The following sections detail a number of parameters associated with the bacterial agents. These parameters are necessarily an abstract representation of the complex underlying mechanisms that contribute to the bacterial cell behaviour. For example, the complex nutrient uptake processes in a cell are reduced to a simple “nutrient intake” parameter. By doing this, we lose

**Table 7.1** Input parameters for bacterial agents in Micro-Gen model

| Input parameter                                                                      | Input value           |
|--------------------------------------------------------------------------------------|-----------------------|
| Biomass threshold for division                                                       | 10,000                |
| Nutrient Intake (b.u. loop <sup>-1</sup> )                                           | 10.0                  |
| Survival cost (b.u. loop <sup>-1</sup> )                                             | 0.2                   |
| Stationary phase relative metabolic rate                                             | 0.2                   |
| Lag phase length (min)                                                               | 66                    |
| $\beta$ -lactamase production rate ( $\mu\text{M s}^{-1}$ )                          | $3.28 \times 10^{-7}$ |
| $\beta$ -lactamase production cost (b.u.)                                            | 0.1                   |
| Antibiotic intake ( $\mu\text{M}$ )                                                  | $6.0 \times 10^{-8}$  |
| Kinetic parameters ( $k_2$ , $K_d$ , $k_{\text{cat}}$ , $K_M$ ),                     | see Table 7.2         |
| Sample values for <i>Staphylococcus aureus</i> species included b.u. = biomass units |                       |

some of the insight that may be gained into the subtle subcellular dynamics of the nutrient uptake mechanism. The emphasis during the design phase was to minimise the number of parameters and aggregate where possible. This results in a simple, lean model that is memory-efficient and thus can be scaled up to very large population sizes even on limited computing hardware.

### 7.2.2.1 Growth Parameters

The biomass of the cell is represented in the simulation by simulation units called “biomass units”. Bacterial agents increase their biomass by absorbing nutrient from the immediate environment. A “nutrient intake” parameter specifies the rate of nutrient absorption by the cell. There is also a “survival cost” associated with normal metabolic activities of the cell, and this is subtracted from the cell biomass in each time step. Reproduction is triggered when the cellular biomass increases beyond a certain threshold for division (“biomass threshold for division”). The cell divides into two identical daughter cells, approximately half the size of the original cell, in a process known as binary fission.

In order to estimate the nutrient intake rate and biomass threshold for division, the model was fitted to an experimentally determined bacterial doubling time of 29 min (generation time of *S. aureus* strain BB255) (Ender et al. 2004). The survival cost parameter influences the length of the stationary phase of the growth cycle. This is modifiable by the user to represent different stationary phase lengths.

A higher survival cost results in a shorter stationary phase because cells enter the death phase more quickly. A survival cost of 2% of the rate of nutrient intake was chosen for the test simulations recorded here. This represents a level which does not limit the exponential phase of bacterial growth. However, in nature this would vary considerably between different strains, and for a more detailed quantitative representation of the growth curve of a particular strain this would need to be estimated from experimental studies.

Another parameter associated with the bacterial agents is the “stationary phase relative metabolic rate”. This parameter is included to take into account the state

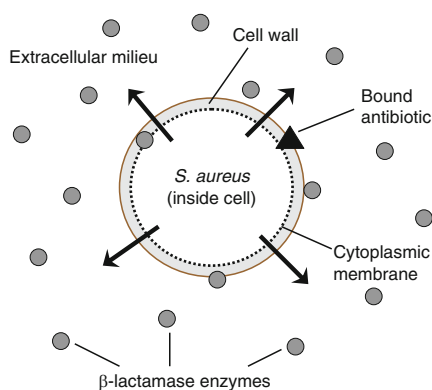
of reduced metabolic activity that bacterial cells enter when they are subjected to severe stress such as nutrient deprivation. The “survival cost” parameter, mentioned above, is reduced to the proportion specified. The principal effect of this parameter is to extend the length of the stationary phase before the bacteria enter the death phase. It represents the bacterium’s ability to preserve itself in hostile, nutrient-deprived conditions by shutting down non-essential metabolic activities. A sample value of 0.2 is used in our test simulations for illustrative purposes; however, as with the “survival cost” parameter, this would need to be experimentally estimated to give a quantitatively accurate representation of the length of the stationary phase.

The “lag phase length” parameter determines the length of time it takes for the bacteria to adapt to their new environment at the start of the simulation. During this phase, bacterial cells adapt to their conditions by synthesising the required cellular components to process the nutrients in their new environment. Their rate of nutrient intake increases until the maximum normal intake rate is achieved. There is a random element introduced by the fact that the bacteria are initialised with different internal nutrient levels at the start of the simulation.

### 7.2.2.2 Antibiotic Resistance Mechanisms

The exposure of bacteria to  $\beta$ -lactam antibiotics triggers the synthesis and release of the  $\beta$ -lactamase enzyme into the extracellular milieu (Fig. 7.4). The  $\beta$ -lactamase production rate is estimated by varying it over a range of values and calculating the minimum inhibitory concentration (MIC) of penicillin G at each value (using experimentally determined kinetic parameters for penicillin G). The MIC is the minimum concentration of an antibiotic that results in complete inhibition of bacterial growth in vitro.

The  $\beta$ -lactamase production rate is calibrated so that the simulated MIC equals the experimentally determined MIC for penicillin G versus the particular bacterial



**Fig. 7.4** Diagram of release of  $\beta$ -lactamase enzymes from *Staphylococcus aureus* cell. Production of  $\beta$ -lactamase is induced by binding of  $\beta$ -lactam antibiotics to a cell surface signal transducer protein (BlaR1). Most of the  $\beta$ -lactamase enzyme is secreted into the extracellular milieu, while some remains bound to the cytoplasmic membrane of the cell

strain being modeled. For the test simulations of MRSA that we have carried out, Type A and Type C  $\beta$ -lactamase-producing MRSA strains were modeled. The experimentally determined MIC results of Norris et al. (1994) for penicillin G (Type A MIC = 72.1  $\mu\text{g/ml}$ ; Type C = 47.9  $\mu\text{g/ml}$ ) were used to estimate the  $\beta$ -lactamase production rates in the simulations.

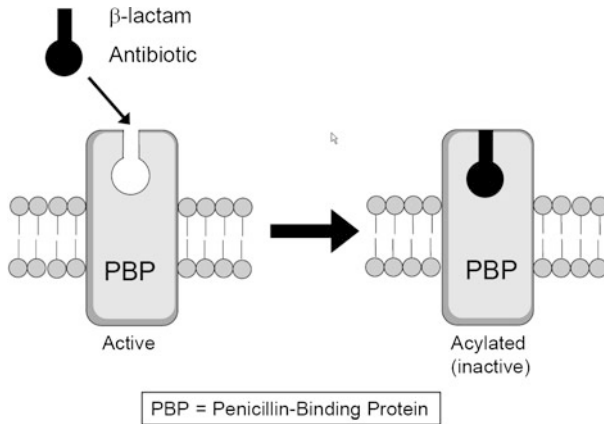
Table 7.1 also lists some parameters for the interactions between the bacterial agent and antibiotic molecules. The “antibiotic intake” parameter determines the rate at which free antibiotic is depleted in the patch by absorption across the cell wall of the bacteria. There are also two kinetic parameters ( $k_2$ ,  $K_d$ ) which determine the rate at which the antibiotic molecules bind to PBPs in the cell membrane. Two kinetic parameters ( $k_{\text{cat}}$ ,  $K_M$ ) are included to describe the interactions between antibiotic molecules and  $\beta$ -lactamase enzymes in the environment. Values for these kinetic parameters were taken directly from experimental literature. A detailed explanation of these kinetic parameters is included in Sects. 7.2.3 and 7.2.4.

### 7.2.2.3 Overcrowding Algorithm

In the case of *S. aureus* bacteria, which are non-motile, an overcrowding algorithm is applied to take into account the physical size constraints of a single patch in the environment. The area of each patch is configured to represent approximately 1  $\mu\text{m}^2$  of medium. An overcrowding algorithm is applied when more bacteria occupy the patch than can be physically accommodated. For example, the estimated cell diameter of a newly divided *S. aureus* cell is 0.5  $\mu\text{m}$  (Giesbrecht et al. 1998). Therefore, when more than four such cells occupy a single patch, the overcrowding algorithm is applied (Fig. 7.3b). The probability,  $p(x_i)$ , of a bacterial cell in an overcrowded patch being moved to an adjacent patch  $i$  is inversely proportional to the relative bacterial biomass in the adjacent patch. The direction a cell is moved is determined by sampling from the resultant probability distribution of the surrounding patches.

### 7.2.3 Antibiotics

The antibiotic level in each patch of the environment is stored as a variable which is subject to diffusion between patches according to Fick’s First Law of diffusion (see Sect. 7.2.1). There is also a half-life associated with the antibiotic, derived from the biological literature, which determines the rate at which the molecule degrades over time (Wishart et al. 2006). Bacterial agents that are in the same patch as the antibiotic will absorb it according to their antibiotic intake rate. The  $\beta$ -lactam antibiotics inhibit bacterial cell division by binding to proteins in the cell membrane called PBPs which are necessary for cell division (Fig. 7.5). If a significant proportion of PBPs in the cell are inactivated, the bacteria will be unable to reproduce and cell death may occur (Giesbrecht et al. 1998).



**Fig. 7.5** Diagram of a  $\beta$ -lactam antibiotic binding to penicillin-binding protein (PBP) in bacterial cell membrane. PBP is an essential component for correct cell wall synthesis during bacterial cell division. However, binding and acylation of the PBP by antibiotic result in inhibition of this function

The simulator includes a quantitative model to estimate the rate of binding of a particular antibiotic to PBPs in the cell membrane, using experimentally calculated kinetic parameters for the reaction. The reaction is a pre-steady state reaction, and therefore the kinetic parameters  $k_2$  (rate of inactivation of PBP2a) and  $K_d$  (dissociation constant) are used to describe it. The ratio of these values ( $k_2/K_d$ ), or the second order rate constant, is a common measure of the antibiotic efficacy at inhibiting PBP function. The proportion of PBP that is inactivated per second,  $k_a$  (the apparent first order rate constant), at a given drug concentration can be calculated as a function of these parameters (7.1) (Chambers et al. 1994).

$$k_a = \frac{k_2[Ab]}{K_d + [Ab]} \quad (7.1)$$

The interactions between the  $\beta$ -lactam antibiotic molecules and PBP2a are explicitly represented in the simulation, but not the interactions with the other PBPs present in the bacterial cell membrane (PBPs 1–4). This is sufficient for representing MRSA because the limiting reaction for antibiotic efficacy is that with PBP2a, which has a binding affinity for  $\beta$ -lactams that is significantly reduced compared to the other PBPs. This minimises the level of complexity caused by introducing more empirical parameters into the model.

Values for the kinetic parameters,  $k_2$  and  $K_d$ , of PBP2a were obtained from experimental studies in the biological literature (Table 7.2) (Fuda et al. 2004; Graves-Woodward and Pratt 1998; Lu et al. 1999). It is possible to estimate the proportion of PBP2a deactivated by antibiotic each time step using these parameters.

**Table 7.2** Experimentally determined kinetic parameters for  $\beta$ -lactam antibiotics versus Type A and Type C  $\beta$ -lactamase-producing MRSA

| Parameter                            | Type A MRSA |         |         | Type C MRSA |         |         |
|--------------------------------------|-------------|---------|---------|-------------|---------|---------|
|                                      | Pen G       | Amp     | Ceph    | Pen G       | Amp     | Ceph    |
| $k_{\text{cat}}$ ( $\text{s}^{-1}$ ) | 171.0       | 308.0   | 0.015   | 210.0       | 355.0   | 0.095   |
| $K_M$ ( $\mu\text{M}$ )              | 51.1        | 255.0   | 6.8     | 55.9        | 122.0   | 5.2     |
| $k_2$ ( $\text{s}^{-1}$ )            | 0.18500     | 0.00470 | 0.00115 | 0.18500     | 0.00470 | 0.00115 |
| $K_d$ ( $\mu\text{M}$ )              | 15,400      | 495     | 586     | 15,400      | 495     | 586     |

*Pen G* Penicillin G, *Amp* Ampicillin, *Ceph* Cephalothin Values derived from biological literature

### 7.2.4 $\beta$ -Lactamase Enzymes

Each bacterial agent has a true/false flag for  $\beta$ -lactamase expression associated with it. When bacterial agents occupy a patch where antibiotic is present,  $\beta$ -lactamase expression is induced (i.e. the flag is changed to true) and there is an exponential increase in the  $\beta$ -lactamase production rate until the maximum rate is reached after approximately 80 min (Lewis et al. 1999). This represents the activation of gene expression mediated by antibiotic binding to the signal-transducer protein BlaR1 and the ensuing time lag during which  $\beta$ -lactamase synthesis is initiated.

The  $\beta$ -lactamase which is then released into the local patch is subject to diffusion according to Fick's First Law of diffusion (Sect. 7.2.1). Also, like the antibiotics, the enzyme has a half-life parameter determining its rate of degradation over time in the environment. If there is no longer any antibiotic left in the bacterium's patch, re-repression of  $\beta$ -lactamase expression occurs and in the model this is represented by the software flag for expression by the bacterial agent being changed to false (corresponding to the BlaR1 protein in nature no longer being auto-activated).

In order to achieve a quantitative representation of the reactions between  $\beta$ -lactamase enzymes and antibiotic molecules, information from kinetic studies is used. The reaction is described using Michaelis–Menten kinetics, with the reaction rate,  $V$ , calculated as the rate at which antibiotic is cleaved (or de-activated) by the enzyme per second (7.2):

$$V = \frac{k_{\text{cat}}[E]_t[Ab]}{K_M + [Ab]} \quad (7.2)$$

There are two principal kinetic parameters used as input to calculate the reaction rate: the turnover rate,  $k_{\text{cat}}$ , and the Michaelis constant,  $K_M$ . The ratio of these values ( $k_{\text{cat}}/K_M$ ) is often used as a measure of enzyme efficiency (Zygmunt et al. 1992). As with the previous kinetic parameters, values for these were derived from the biological literature (Table 7.2) (Zygmunt et al. 1992).  $[Ab]$  and  $[E]_t$  refer to the concentrations of antibiotic  $\beta$ -lactamase enzyme (sum of both free and occupied enzyme) in the local patch, respectively.

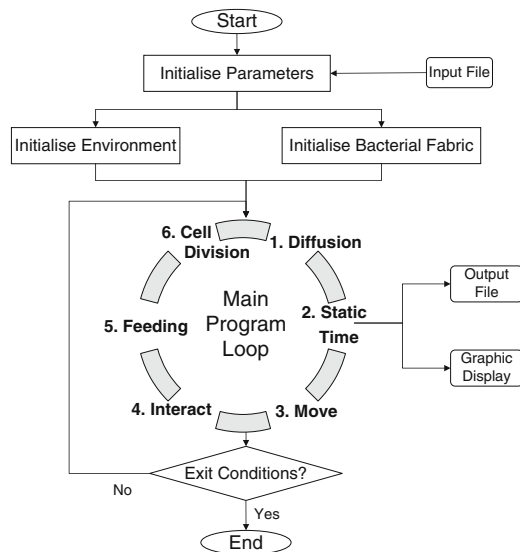
### 7.2.5 Program Flow Structure

The structure of the program consists of an initialisation stage during which the main components of the environment, the patches and the bacterial agents are initialised. This is followed by the principal program loop during which the main tasks of the simulation are carried out in successive discrete time steps until certain exit conditions (e.g. no bacteria left) are reached (Fig. 7.6). The loop is configured to represent approximately 2 s of real-time, although this is modifiable by the user to apply a different temporal granularity. Time-dependent input parameters, such as the nutrient intake rate or antibiotic kinetic parameters, are then configured to the specified timescale.

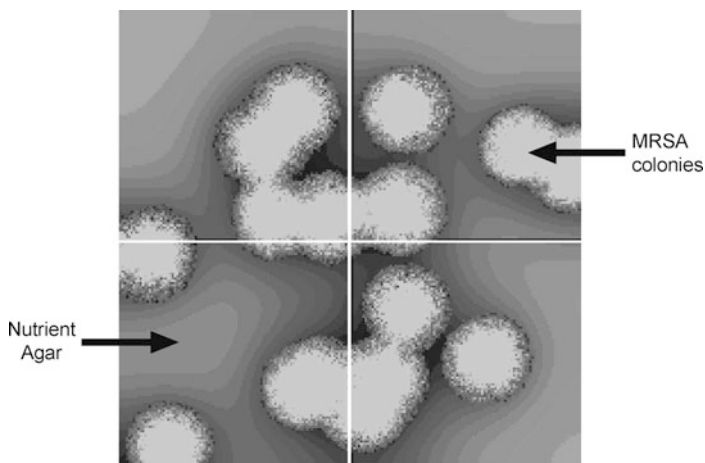
There are six stages in the program loop during which the activities of the simulation are carried out. Stage 1 (diffusion) involves updating the levels of the various molecular components (nutrients, antibiotics and enzymes) in the environment. The movement of molecules between patches is calculated by applying the algorithm for Fick's First Law of diffusion.

Stage 2 (static time) consists of several housekeeping processes. These include subtracting a survival cost from the bacterium's energy stock, representing energy expended on basic metabolic processes in the cell. If  $\beta$ -lactamase gene expression is active, then the enzyme is released into the environment, with an associated energy cost. There is also an optional graphical output that the program can produce to display the simulated culture environment and the levels of various constituents in it (Fig. 7.7).

A movement/overcrowding algorithm can be applied in stage 3 of the loop to update the positions of the bacterial agents. This comes into effect when the density



**Fig. 7.6** Diagram of program flow structure in Micro-Gen Bacteria Simulator with principal stages of program loop labelled



**Fig. 7.7** Screenshot of *Micro-Gen* simulation running in parallel on four computing nodes. Circular shaped MRSA bacterial colonies can be observed growing on a background of nutrient medium. Contours produced by different shades of grey represent nutrient gradient (lighter colour equals higher nutrient concentration)

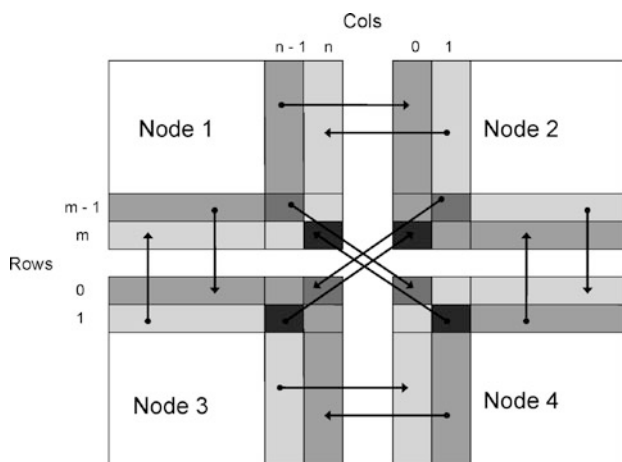
of cells in a patch exceeds the size limit of the patch, and thus there is a possibility of one or more of the cells being moved into an adjacent patch (see Sect. 7.2.2.3).

Stage 4 is when the algorithms defining the interactions between antibiotics and bacterial cells and/or  $\beta$ -lactamase enzymes are applied. Kinetic parameters derived from biological studies are used to determine the reaction rates (7.1), (7.2). Stage 5 of the loop (feeding) is when the bacterial agents can absorb nutrient from the environment and increase their internal energy stocks. In the case of the simulations carried out for this study, the rate of intake is that which will result in a generation time of 29 min (Ender et al. 2004).

The final stage of the loop is when binary fission occurs, whereby the bacterial agents reproduce asexually to produce two identical daughter cells. This can occur if the bacterial cell's energy stock has exceeded a certain threshold for division, and the level of antibiotic damage (measured by the proportion of inactivated PBP) is below a critical level.

## 7.2.6 Parallelisation

The agent-based approach to modeling bacterial colonies means that the computational resources increase proportionately with the number of agents required. Therefore, to scale simulations up to represent very large population sizes, it is important to be able to take advantage of modern high performance computing resources. The size of the bacterial populations in nature range from  $10^6$  to  $10^{10}$  cells per millilitre of seawater with even higher concentration in individuals



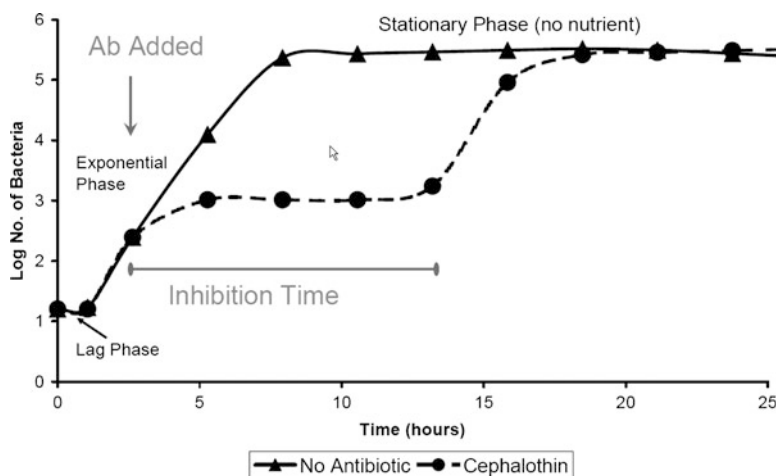
**Fig. 7.8** Diagram of communication between adjacent nodes at overlapping boundary conditions when Micro-Gen is run in a parallel configuration

suffering from acute infections ([Guan and Kamino 2001](#)). Therefore, Micro-Gen was designed to take advantage of parallel computing resources and achieve an efficient scale up to hundreds and even potentially thousands of computing processors.

It does this by dividing the simulated environment equally among all available computing nodes so that each processor is responsible for computing only a subsection of the overall population ([Fig. 7.7](#)). In cases where bacteria and/or molecules move across the boundaries separating two nodes, a communication strategy is used whereby overlapping boundary conditions exist between nodes where the elements are exchanged ([Fig. 7.8](#)). They are sent to the receiving node across the available network connection using a communication protocol known as the message passing interface (MPI).

### 7.3 Simulations of Bacteria–Antibiotic Interactions

The complex relationships between the biomolecular/kinetic properties of drug compounds and emergent pharmacodynamic parameters, such as the MIC, are an important area to study to develop optimal drug treatment regimens and for better rational drug design strategies ([Regoes et al. 2004](#)). It provides a basis for understanding the dynamics involved in the development of antibiotic resistance and thus can lead to better strategies for limiting its expansion. Micro-Gen represents a good theoretical framework for analysing the *in vitro* dynamics of antibiotics interacting with bacteria and for connecting knowledge from low-level biochemical studies with emergent properties of the population. Furthermore, future work will involve extending the model to represent the more complex dynamics found in the *in vivo* clinical setting as well.



**Fig. 7.9** Effect of antibiotic exposure on simulated growth curve of MRSA bacteria in nutrient-limited culture conditions. Cephalothin antibiotic (103.1  $\mu\text{g/ml}$ ) added after 3.5 h of incubation, during the exponential phase of colony growth. This results in inhibition of colony growth for a finite period of time (inhibition time) until depletion of antibiotic according to its natural half-life, or hydrolysis by enzymes such as  $\beta$ -lactamases, allows growth to resume

In order to demonstrate the capabilities of the Micro-Gen model, some sample results are included in this section from simulations of MRSA bacteria growing in batch culture. Figure 7.9 shows the simulated growth curve of a sample MRSA bacterial colony outputted by Micro-Gen. The control culture of MRSA, where no antibiotic is added, follows the characteristic standard growth curve of bacteria grown in nutrient-limited culture conditions (lag, log and stationary phases – death phase not shown). A second curve shows the effect of adding an inhibitory dose of antibiotic (103.1  $\mu\text{g/ml}$  of cephalothin), after 3.5 h of incubation, in the development of the colony.

The addition of an inhibitory dose of antibiotic during the exponential phase of growth causes inhibition of growth for a finite period of time. The length of time bacterial growth is inhibited is important as it determines the recommended dosage regimen for an antibiotic. Growth must be inhibited for a long enough period of time to cover the gap between successive doses of medication to effectively block bacterial expansion, so that the immune response can remove the infection. Factors such as the half-life of the antibiotic and the action of bacterial enzymes, such as  $\beta$ -lactamases which degrade the antibiotic molecules, influence the length of the inhibition time.

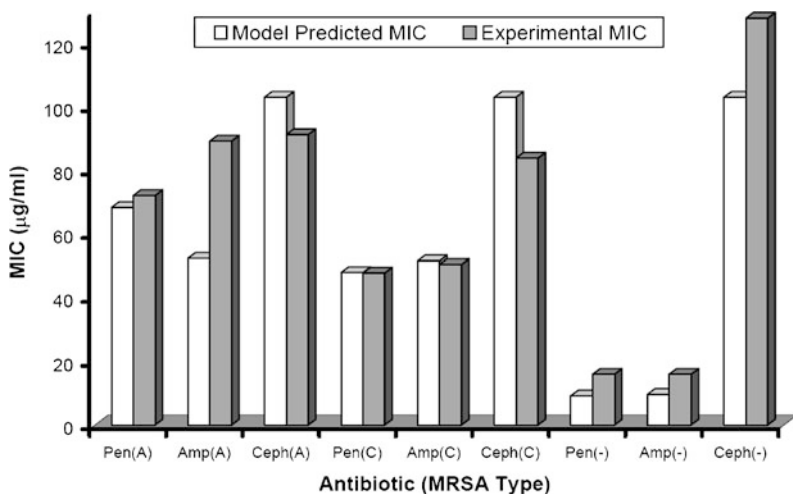
Some test simulations were carried out to validate the model and verify that the algorithm representing bacteria–antibiotic interactions reproduces the effects observed in real-life experiments (Murphy et al. 2008). In order to do this, parameters from the biological literature applicable to three types of MRSA bacteria were used for the test simulations. These types of MRSA are differentiated by their

$\beta$ -lactamase status: Type A MRSA and Type C MRSA are named because they produce  $\beta$ -lactamase enzymes of these respective types with different kinetic parameters associated with them. The third type tested was a  $\beta$ -lactamase-negative strain which was included as a control.

Type A and Type C  $\beta$ -lactamase enzymes are distinguished by their kinetic parameters ( $k_{cat}/K_M$ ), and values for these were derived from experimental literature (see Table 7.2). They were chosen for this study because they are the most common types of  $\beta$ -lactamase found in MRSA bacteria. A study by Norris et al. (1994) found that among 50  $\beta$ -lactamase-producing MRSA isolates taken from nine locations across the USA, 80% expressed Type A  $\beta$ -lactamase and the remainder expressed Type C. Type B and Type D  $\beta$ -lactamases are thought to be less common among MRSA strains.

The MIC was calculated from the model for a number of common  $\beta$ -lactam antibiotics against the MRSA strains, and compared with the real-world results. The MICs were estimated from the model in an analogous way to the broth dilution test carried out in experimental studies: a series of simulations were performed with stepwise increases of the concentration of antibiotic in each run. The recorded MIC is the minimum concentration of antibiotic that resulted in complete inhibition of bacterial growth for a pre-determined length of time.

Figure 7.10 shows the predicted MIC values from Micro-Gen for three common  $\beta$ -lactam antibiotics (penicillin G, ampicillin and cephalothin) compared



**Fig. 7.10** Predicted (Micro-Gen) versus experimentally determined MICs ( $\mu\text{g/ml}$ ) of penicillin G, ampicillin and cephalothin antibiotics against three different types of MRSA. Experimentally determined MICs for  $\beta$ -lactamase-positive (Types A and C) strains are from Norris et al. (1994). Experimentally determined MICs for  $\beta$ -lactamase-negative strains are from Malouin et al. (2003). Predicted MICs are derived from triplicate simulations with the geometric mean MIC  $\pm$  SEM ( $\mu\text{g/ml}$ ) displayed. (A) Type A  $\beta$ -lactamase producing MRSA, (C) Type C  $\beta$ -lactamase producing MRSA, (–)  $\beta$ -lactamase-negative MRSA

with results from the experimental studies published in the scientific literature (Murphy et al. 2008; Malouin et al. 2003; Norris et al. 1994). There is clearly a very good correspondence between the predicted values from Micro-Gen and the real-world situation, despite the fact that parameters from different experimental studies carried out in different laboratories were used. This indicates that despite the minor differences that arise from varying experimental conditions, the overall MICs of the antibiotics are primarily determined by their low-level kinetic parameters.

The comparison between the predicted and experimental results is a useful initial step for validating the model. However, this only confirms existing knowledge about these particular antibiotics. The true power of the model exists in being able to deconstruct these values and relate them back to the low-level molecular and cellular parameters of the individual bacterial cells. Through varying the individual parameters, one can gain a better understanding of the root causes of the observed drug resistance. For example, resistance of MRSA to cephalothin is mediated by the antibiotic's inability to bind efficiently to the PBP2a protein on the bacterial cell surface ( $k_2/K_d = 1.96 \text{ M}^{-1} \text{ s}^{-1}$ ). By contrast, penicillin G binds more readily to PBP2a ( $k_2/K_d = 12.0 \text{ M}^{-1} \text{ s}^{-1}$ ), but its higher susceptibility to cleavage by  $\beta$ -lactamase enzyme negates this advantage. This is a simple example of how resistance can be the result of different underlying mechanisms and the dynamic interplay between them. The modeling approach allows us to investigate a wide range of possible scenarios to investigate the underlying mechanistic basis for the resistance.

These results were obtained without attempting to “fit” the kinetic input parameters to the MIC results. The only parameter, which influences the kinetics of the bacterial agents, that required to be fitted was the  $\beta$ -lactamase production rate. However, for the results of the  $\beta$ -lactamase-negative strain in Fig. 7.10, this fitting step did not have to be carried out. Therefore, the predicted MICs were purely an emergent property of the inputted kinetic parameters for the PBP2a-antibiotic binding reaction. There was still a close quantitative agreement between the experimental and simulation results even in this scenario.

We have carried out extensive research and sensitivity analyses to examine the impacts of the various molecular/environmental factors that lead to the observed responses to antibiotic treatment. These have included explorations of the parameter space with particular emphasis on the kinetic parameters associated with the antibiotic molecules. A detailed discussion of these studies is beyond the scope of this chapter; however, for further information the reader is directed towards a couple of significant publications on the subject (Murphy et al. 2008, 2009). In general, the reaction of a bacterial community to antibiotic treatment is sensitive to environmental conditions such as the rate of diffusion, temperature, pH, etc. These factors can play an important role in contributing to the differences observed between results in a laboratory setting and the actual in vivo response to treatment. Simulations are useful for examining situations not readily reproduced in the experimental setting due to logistical constraints. For example, the model can be used to examine evolutionary pathways and assess the fitness advantages/disadvantages associated with, for example, changes in the reaction profiles of the bacteria–antibiotic interactions.

Knowledge from these studies would be useful in rational drug design to assess the potential for antibiotic resistance to develop rapidly following introduction of a drug into clinical use.

## 7.4 Conclusions and Future Work

An integrated systems-level understanding of the complex dynamics that lead to the rapid development and spread of antibiotic resistance within bacterial populations is required to meet the growing demands of treating patients with multi-drug resistant bacterial infections. In order to meet this challenge, scientists will need to approach the problem from various angles in both the theoretical and experimental domains. In terms of theoretical approaches to understanding the system dynamics of resistance development, there are a number of techniques available, some of which were mentioned at the start of this chapter, such as ODEs and CA. However, the agent-based approach discussed here is a particularly important tool for investigating the relationship between the individual molecular components and bacterial cells of the system and the overall treatment outcome.

There is much scope for improvement and development of the agent-based approach as it is only relatively recently that it has begun to gain traction as an important addition to the bacterial population modeling toolset. Its potential applications are extremely broad, and it represents a great means for investigating complex situations such as the diverse, heterogeneous bacterial ecosystems present in the intestinal tract or the network of genetic exchanges arising from horizontal gene transfer in bacterial populations (Sorensen et al. 2005). The modular nature of the agent-based approach and its “bottom-up” approach to modeling systems means that seemingly intractable problems may be broken down to their simplest, basic units and new insights gained into the emergence of complexity from their interactions.

Future work will include using the model to investigate the system dynamics of combination therapy where multiple classes of antibiotic are applied simultaneously to treat infections. It can also be used to test hypothetical scenarios by varying the parameters of existing antibiotics to explore how potential novel compounds might act. The agent-based approach is also suitable for modeling the evolution of antibiotic resistance over time by incorporating genetic components into the bacterial agents. This would allow the examination of both the temporal and spatial dynamics of antibiotic resistance development in a population exposed to antibiotics.

Other important developments of the model include expanding the environment to represent three-dimensional space to model more complex spatially structured microbial communities such as biofilms. Biofilms are complex aggregations of microbial cells that are characterised by their complex cellular interactions, genetic diversity and structural heterogeneity. They are characterised by highly heterogeneous localised niches where the chemistry varies dramatically over small distances. The agent-based approach is a powerful tool for modeling the interactions within a heterogeneous environment such as this.

**Acknowledgements** The authors would like to thank Marc Devocelle from the Royal College of Surgeons in Ireland, who collaborated on the biological aspects of the Micro-Gen research project. The authors would also like to acknowledge the contributions of Mathieu Joubert, Grainne Kerr, Chris Pender, Marian Duggan and Ronan Winters who developed the original BAIT software tool under the supervision of Ray Walshe in Dublin City University.

## References

- E. P. Abraham and E. Chain. An enzyme from bacteria able to destroy penicillin. *Nature*, 146:837, 1940
- P. G. Ambrose, S. M. Bhavnani, C. M. Rubino, A. Louie, T. Gumbo, A. Forrest, and G. L. Drusano. Pharmacokinetics-pharmacodynamics of antimicrobial therapy: it's not just for mice anymore. *Clinical Infectious Diseases*, 44(1):79–86, 2007
- E. Ben-Jacob, O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, and T. Vicsek. Generic modelling of cooperative growth patterns in bacterial colonies. *Nature*, 368(6466):46–49, 1994
- A. Bondi and C. C. Dietz. Penicillin resistant staphylococci. *Proceedings of the Society for Experimental Biology and Medicine*, 60(1):55–58, 1945
- E. Chain, H. W. Florey, M. B. Adelaide, A. D. Gardner, N. G. Heatley, M. A. Jennings, J. Orr-Ewing, and A. G. Sanders. Penicillin as a chemotherapeutic agent. 1940. *Clinical Orthopaedics and Related Research*, 295:3–7, 1993
- H. F. Chambers, M. J. Sachdeva, and C. J. Hackbarth. Kinetics of penicillin binding to penicillin-binding proteins of *Staphylococcus aureus*. *The Biochemical Journal*, 301(Pt 1):139–144, 1994
- D. A. Devine. Cationic antimicrobial peptides in regulation of commensal and pathogenic microbial populations. *Mammalian Host Defense Peptides*, pages 9–39, 2004
- A. Elfving, Y. LeMarc, J. Baranyi, and A. Ballagi. Observing growth and division of large numbers of individual bacteria by image analysis. *Applied and Environmental Microbiology*, 70(2): 675–678, 2004
- M. Ender, N. McCallum, R. Adhikari, and B. Berger-Bachi. Fitness cost of secmec and methicillin resistance levels in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 48(6): 2295–2297, 2004
- K. R. Eriksen. Celbenin-resistant staphylococci. *Ugeskrift for laeger*, 123:384–386, 1961
- J. F. Fisher, S. O. Meroueh, and S. Mobashery. Bacterial resistance to beta-lactam antibiotics: compelling opportunism, compelling opportunity. *Chemical Reviews*, 105(2):395–424, 2005
- C. Fuda, M. Suvorov, S. B. Vakulenko, and S. Mobashery. The basis for resistance to beta-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *The Journal of Biological Chemistry*, 279(39):40802–40806, 2004
- C. C. Fuda, J. F. Fisher, and S. Mobashery. Beta-lactam resistance in *Staphylococcus aureus*: the adaptive resistance of a plastic genome. *Cellular and Molecular Life Sciences*, 62(22): 2617–2633, 2005
- M. Gardner. Mathematical games: The fantastic combinations of john conways new solitaire game life. *Scientific American*, 223(4):120–123, 1970
- P. Giesbrecht, T. Kersten, H. Maidhof, and J. Wecke. Staphylococcal cell wall: morphogenesis and fatal variations in the presence of penicillin. *Microbiology and Molecular Biology Reviews*, 62 (4):1371–1414, 1998
- M. Ginovart, D. Lopez, and J. Valls. Indisim, an individual-based discrete simulation model to study bacterial cultures. *Journal of Theoretical Biology*, 214(2):305–319, 2002
- K. Graves-Woodward and R. F. Pratt. Reaction of soluble penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus* with beta-lactams and acyclic substrates: kinetics in homogeneous solution. *The Biochemical Journal*, 332(3):755–761, 1998
- V. Grimm. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological modelling*, 115(2–3):129–148, 1999

- M. J. Grimson and G. C. Barker. Continuum model for the spatiotemporal growth of bacterial colonies. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 49(2):1680–1684, 1994
- L. L. Guan and K. Kamino. Bacterial response to siderophore and quorum-sensing chemical signals in the seawater microbial community. *BMC Microbiology*, 1:27, 2001
- N. R. Jennings, K. Sycara, and M. J. Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38, 1998
- W. M. M. Kirby. Extraction of a highly potent penicillin inactivator from penicillin resistant staphylococci. *Science*, 99(2579):452–453, 1944
- J. U. Kreft, G. Booth, and J. W. Wimpenny. Bacsim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*, 144(12):3275–3287, 1998
- A. M. Lacasta, I. R. Cantalapiedra, C. E. Auguet, A. Penaranda, and L. Ramirez-Piscina. Modeling of spatiotemporal patterns in bacterial colonies. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 59(6):7036–7041, 1999
- S. B. Levy and B. Marshall. Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine*, 10(12 Suppl):S122–129, 2004
- R. A. Lewis, S. P. Curnock, and K. G. Dyke. Proteolytic cleavage of the repressor (blai) of beta-lactamase synthesis in *Staphylococcus aureus*. *FEMS Microbiology Letters*, 178(2):271–275, 1999
- W. P. Lu, Y. Sun, M. D. Bauer, S. Paule, P. M. Koenigs, and W. G. Kraft. Penicillin-binding protein 2a from methicillin-resistant *Staphylococcus aureus*: kinetic characterization of its interactions with beta-lactams using electrospray mass spectrometry. *Biochemistry*, 38(20):6537–6546, 1999
- F. Malouin, J. Blais, S. Chamberland, M. Hoang, C. Park, C. Chan, K. Mathias, S. Hakem, K. Dupree, E. Liu, T. Nguyen, and M. N. Dudley. Rwj-54428 (mc-02,479), a new cephalosporin with high affinity for penicillin-binding proteins, including pbp 2a, and stability to staphylococcal beta-lactamases. *Antimicrobial Agents and Chemotherapy*, 47(2):658–664, 2003
- W. Materi and D. S. Wishart. Computational systems biology in drug discovery and development: methods and applications. *Drug Discovery Today*, 12(7–8):295–303, 2007
- J. T. Murphy and R. Walshe. Micro-gen: an agent-based model of bacteria-antibiotic interactions in batch culture. In *Proceedings of the 20th Annual European Simulation and Modelling Conference*, pages 239–242. Eurosis-ETI, October 23–25, 2006
- J. T. Murphy, R. Walshe, and Marc Devocelle. Agent-based model of methicillin-resistant *Staphylococcus aureus* and antibiotics in batch culture. In *Proceedings of 21st Annual European Simulation and Modelling Conference*, pages 409–414. Eurosis-ETI, October 20–22, 2007
- J. T. Murphy, R. Walshe, and M. Devocelle. A computational model of antibiotic-resistance mechanisms in methicillin-resistant *Staphylococcus aureus* (MRSA). *Journal of Theoretical Biology*, 254(2):284–293, 2008
- J. T. Murphy, R. Walshe, and M. Devocelle. Modelling the population dynamics of antibiotic-resistant bacteria: An agent-based approach. *International Journal of Modern Physics C*, 20(3):435–457, 2009
- H. C. Neu. The crisis in antibiotic resistance. *Science*, 257(5073):1064–1073, 1992
- J. B. Nielsen and J. O. Lampen. Membrane-bound penicillinases in gram-positive bacteria. *The Journal of Biological Chemistry*, 257(8):4490–4495, 1982
- G. W. Niven, T. Fuks, J. S. Morton, S. A. Rua, and B. M. Mackey. A novel method for measuring lag times in division of individual bacterial cells using image analysis. *Journal of Microbiological Methods*, 65(2):311–317, 2006
- S. R. Norris, C. W. Stratton, and D. S. Kernodle. Production of a and c variants of staphylococcal beta-lactamase by methicillin-resistant strains of *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 38(7):1649–1650, 1994
- R. R. Regoes, C. Wiuff, R. M. Zappala, K. N. Garner, F. Baquero, and B. R. Levin. Pharmacodynamic functions: a multiparameter approach to the design of antibiotic treatment regimens. *Antimicrobial Agents and Chemotherapy*, 48(10):3670–3676, 2004

- D. Skinner and C. S. Keefer. Significance of bacteremia caused by *Staphylococcus aureus*. *Archives of Internal Medicine*, 68:851–875, 1941
- S. J. Sorensen, M. Bailey, L. H. Hansen, N. Kroer, and S. Wuertz. Studying plasmid horizontal transfer in situ: a critical review. *Nature Reviews Microbiology*, 3(9):700–710, 2005
- M. W. van der Woude and A. J. Baumler. Phase and antigenic variation in bacteria. *Clinical microbiology reviews*, 17(3):581–611, 2004
- C. Walsh. Molecular mechanisms that confer antibacterial drug resistance. *Nature*, 406(6797):775–781, 2000
- R. Walshe. Modelling bacterial growth patterns in the presence of antibiotic. In *Proceedings of the 11th IEEE International Conference on Engineering of Complex Computer Systems*, pages 177–188, Washington, DC, USA, 15–17 August 2006. IEEE Computer Society
- D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(Database issue):D668–672, 2006
- G. D. Wright. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Reviews Microbiology*, 5(3):175–186, 2007
- H. Z. Zhang, C. J. Hackbarth, K. M. Chansky, and H. F. Chambers. A proteolytic transmembrane signaling pathway and resistance to beta-lactams in staphylococci. *Science*, 291(5510):1962–1965, 2001
- D. J. Zygmunt, C. W. Stratton, and D. S. Kernodle. Characterization of four beta-lactamases produced by *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 36(2):440–445, 1992

# Chapter 8

## Modeling the Spatial Pattern Forming Modules in Mitotic Spindle Assembly

Chaitanya A. Athale

### 8.1 Introduction

The phyllotaxy of leaves, the scaling of animal limbs and organs, the cellular patterns in tissue, the shapes of individual cells, the form of subcellular structures like spindles, cilia and flagella as well as polarization of signalling molecules all encompass biological pattern formation. Many of these are caused by underlying processes that are in common with physical processes that generate nonliving patterns – crystals, faults, river-networks and chemical oscillators. This process involves both self-assembly and self-organization. While self-assembly is driven to energy minima attained by interacting components and often has rather deterministic ends, self-organization is driven often by local interactions with feedback which lead to unpredictable outcomes.

The most prominent example of modeling biological spatial pattern formation was the use of reaction-diffusion models by Alan Turing to explain the appearance of periodic concentrations of a chemical that would generate periodic morphogenetic patterns. Using a combination of a slow-diffusing activator and a fast diffusing inhibitor, Turing could show that given an initial random distribution of both species, the system within a range of reaction parameters would produce a steady-state periodic pattern. This counter-intuitive finding about the emergence of order from a random distribution was explained by the mathematical properties of such a system. At the moment experimental evidence of similar Turing patterns has been demonstrated in few systems like the *Escherichia coli* oscillating Min proteins which help find the centre of the bacterium during division (Meinhardt and de Boer 2001). This model however has more components and different interactions from a minimal activator substrate-depletion system which could theoretically produce the same patterns. Some of the complexity of the network is necessary to

---

C.A. Athale (✉)

EMBL, Meyerhofstrasse 1, Heidelberg 69117, Germany

and

IISER Pune, IISER, Central Tower, Sai Trinity Building, Sutarwadi Road,

Pashan, Pune 411021, India

e-mail: [athale@embl.de](mailto:athale@embl.de); [cathale@iiserpune.ac.in](mailto:cathale@iiserpune.ac.in)

robustly suppress certain modes of the system, demonstrating that pattern forming reaction-diffusion systems in biology do not necessarily show the minimal activator-inhibitor system although similar principles of diffusion driven instability might govern them.

Spindle assembly occurs in eukaryotic cells in the M-phase of cell division. In the stage between prophase and metaphase, simultaneously the chromatin condenses, the nuclear envelope ruptures, a pair of centrosomes move to opposite ends of the nucleus and microtubule fibers nucleated from the two poles attach to chromosomes as well as the opposite poles to form a mitotic spindle. In some cells (e.g., yeast) spindle assembly is not accompanied by nuclear envelope breakdown, while in others such as starfish oocytes, congression of chromosomes is driven by acto-myosin dependent chromosome motility to the metaphase plate. In meiosis for instance, the spindle lack specific centrosomes, instead the poles self-organize by a combination of chromosomal nucleation and minus-end directed transport of microtubules. Taken together the mitotic spindle is a machine that is responsible for segregation of the chromosomes and its assembly is of fundamental importance for inheritance, cancer and a fundamental understanding of biological self-replication. Its assembly can be seen as a process of pattern formation involving different modules and processes. Although the discovery of the mitotic spindle is over 120 years old dating to work by [Flemming \(1882\)](#), and many of the motor proteins (as reviewed in [Walczak et al. \(1998\)](#)) and the network of regulators (reviewed in [Akhmanova and Steinmetz \(2008\)](#)) have been found, we still do not understand how these parts function together to produce the complex machinery of the spindle. Such understanding will not just have consequences for the field of cell division but might also act as a template for solving other problems in pattern-forming systems at the cellular scale.

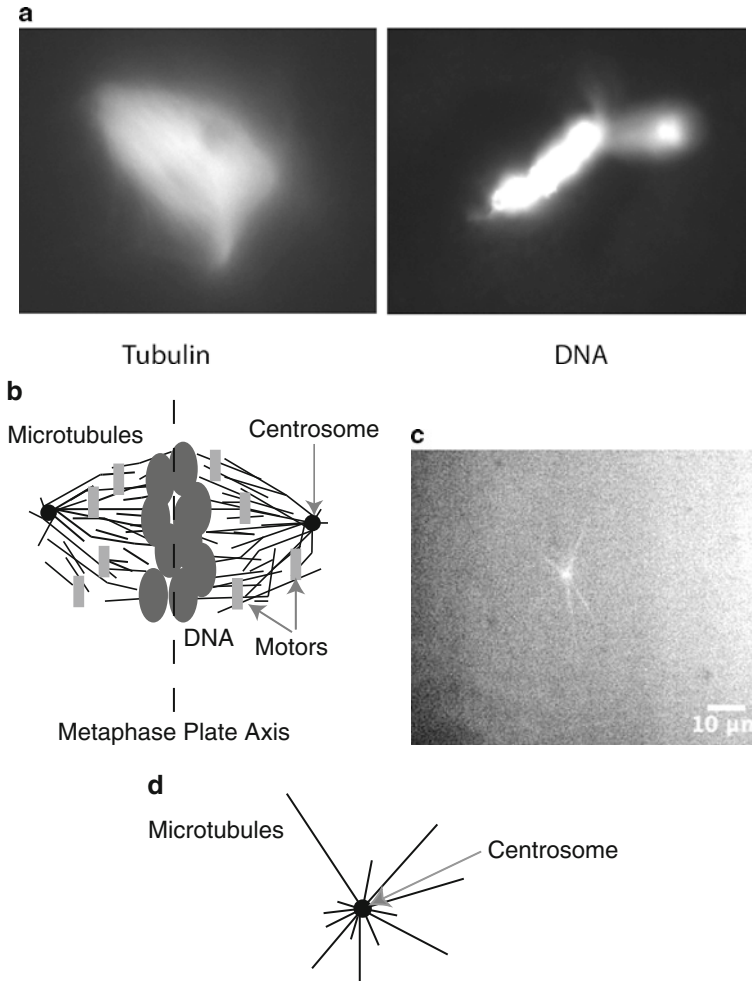
Applying a systems biology approach that combines experiment reconciled to theory promises to reveal some of the aspects of this complex pattern forming system. Modeling has been successfully applied to understanding the assembly of the mitotic spindle. These models have approached this complex problem by reducing the spindle assembly process into modules. These modules can be classified as:

- Microtubule dynamics
- Microtubule-motor interactions
- Chromosome dynamics
- Reaction-diffusion gradients of microtubule dynamics regulation

Spindle-assembly is a complex combination of various processes like reaction-diffusion, mechanical contact, nonlinear interactions and spatial-pattern formation. In order to understand it, even as a model, it is necessary to study the modules that govern it. In recent work, this author and co-workers have taken such an approach, separating the effect of the reaction-diffusion components from the mechanical in spindle assembly, and in the process discovered novel subcellular gradient shapes ([Athale et al. 2008](#)).

Experiments of spindle assembly have been performed in several model systems. The *minimal system* is the commonly used biochemically pure in vitro system with microtubules, motors, *microtubule associated proteins (MAPs)* and DNA. It has

not yet been possible to assemble a fully functional spindle in such a reduced system, though aspects like the localization of proteins to microtubule tips have been recently successfully reproduced (Bieling et al. 2007). The use of cytoplasmic extracts from the eggs of the African clawed frog *Xenopus laevis* with fluorescently labelled tubulin and DNA to examine the mitotic spindle (Fig. 8.1a, b) has



**Fig. 8.1** *Fluorescence microscopy and schematic view of spindle.* (a) A fluorescence microscopy image of the same spindle assembled in *Xenopus* egg extracts and fixed on a slide with the Cy3 (a cyanine fluorescent dye) tubulin labelled microtubules (left) and Hoechst labelled sperm DNA (right) imaged on a wide-field microscope. (b) A schematic view of the spindle depicts the major components: centrosomes, microtubules chromosomal DNA and microtubule-dependent motors. (c) The centrosome when placed in mitotic *Xenopus* egg extract with fluorescent tubulin nucleates a star-like distribution of microtubules, referred to as the centrosomal aster. (d) A schematic view displays the two major components of an aster, the centrosome which nucleates microtubules and the radiating microtubules

become widespread. Simply placing one of the components, the centrosome in the extract also produces dynamic microtubule asters (Fig. 8.1c, d). Spindle assembly has also been demonstrated using plasmid DNA in the *Xenopus* egg-extract. This DNA lacks any chromosome-specific information and even centrosomes are not required to assemble the spindle (Heald et al. 1996). Spindle assembly appears thus to be self-organized and has many redundant pathways. Another commonly used system to study spindle assembly is fission yeast *Schizosaccharomyces pombe* that combines easy genetic and mechanical manipulation of the spindle. A more complex context for spindle assembly is the *Drosophila* syncytial blastoderm stage embryo (cycles 10–13) with up to  $10^3$  nuclei undergoing synchronized mitoses. How these modules of spindle assembly have been used to study spindle assembly by experimental and theoretical approaches will be further elaborated upon in the following sections.

## 8.2 Microtubule Dynamics

The microtubule cytoskeleton is a critical structural component in the process of assembling the mitotic spindle. This protein forms an oriented polymer showing a dynamic *plus-end* that are mostly found to grow, while *minus-ends* are those that mostly disassemble by losing subunits (Fig. 8.2a, b). The assembly of the spindle occurs typically within 20–30 min (Wollman et al. 2005) and requires dynamic changes in its components. This is seen in the case of the microtubule cytoskeleton when it transitions from its interphasic to mitotic state. The interphasic microtubules extend from the centrosome to cell periphery and show little change in length. Mitotic microtubules in contrast are shorter, more dynamic and appearing to randomly switch between growing and shrinking states (Verde et al. 1992). What triggers this transition from interphasic to mitotic microtubules dynamics is still a subject of active research. Experimental findings suggest that discrete phosphorylation states of the network of MAPS may play a role in this transition (Niethammer et al. 2007). Microtubules can be broadly defined to have two main aspects – *nucleation* and *polymerization*.

### 8.2.1 Nucleation

The nucleation of microtubules *in vivo* occurs most commonly from centrosomes which act as templates for tubules to form. Verde et al. (1992) modeled the centrosomal nucleation rate as a constant. However, a later study modeled the nucleation of microtubules from centrosomes using a mean field approach. The study demonstrated the existence of two regimes – one that is nucleation “site limited” for small number of nucleation sites and another that is “diffusion limited” for large numbers of nucleation sites (Dogterom et al. 1995). The diffusion limitation arises

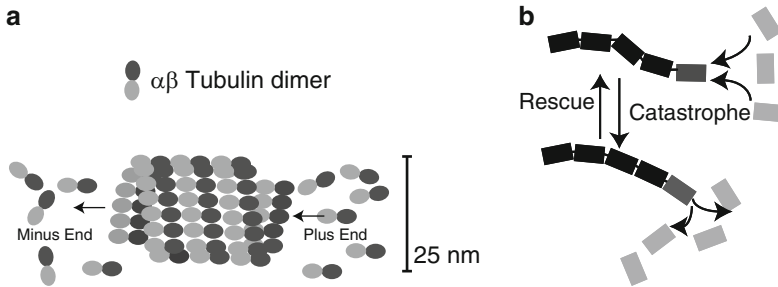
from a large numbers of nucleation sites that consume GTP-tubulin locally around centrosomes due to polymerization. Such physical mechanisms are in contrast to conventionally accepted biochemical regulation of nucleation (Raynaud-Messina and Merdes 2007). In addition plant cells and meiotic spindles do not require the presence of centrosomes to nucleate microtubules. Instead other nucleation centres at the cortex (Ehrhardt 2008) and chromosomes (Gruss et al. 2001) locally nucleate microtubules. This module of non-centrosomal nucleation has been used to model spindle assembly by in a “slide and cluster” model by combining motor proteins with microtubule nucleation on existing microtubules (Burbank et al. 2007; Clausen and Ribbeck 2007).

### 8.2.2 Polymerization

Microtubule polymerization proceeds by random transition of between growing and shrinking states that has been defined by Mitchison and Kirschner (1984) as *dynamic instability*. They further proposed that the experimental dynamics of mitotic microtubules indicate out of equilibrium oscillations due to stabilization/destabilization in the presence/absence of a GTP capped state. Simultaneously a general model of a two-state polymer was developed that could explain the trends in experiment (Hill 1984). Monte Carlo simulations of this two-state polymer model were used to explore the parameter space that reproduced the experimental microtubule length change (Chen and Hill 1985). The transition of a growing polymer to shrinking state was defined as *catastrophe* and the transition from shrinking to growing defined as *rescue*. More recent work on the transition of microtubule dynamics between interphase and mitosis modeled microtubule dynamics of polymerization by dynamical probabilistic equations (Verde et al. 1992). The model parameters are – frequencies of rescue ( $f_{\text{res}}$ ) and catastrophe ( $f_{\text{cat}}$ ), and velocities of growth ( $v_g$ ) and shrinkage ( $v_s$ ). Catastrophe refers to the transition from growing to shrinking state and rescue is the transition from shrinking to growing state. The average length ( $\langle L \rangle$ ) of the polymer thus depends on the flux and is given by:

$$\langle L \rangle = \frac{v_g v_s}{v_s f_{\text{cat}} - v_g f_{\text{res}}}. \quad (8.1)$$

Such a model simplifies the known structural components of microtubule fibers that consist of 12–14 protofilaments of dimeric  $\alpha$  and  $\beta$  tubulin organized in a spiral (Fig. 8.2a). Additionally one end of this tube grows faster than the other – referred to as the plus-end. The plus-end is thought originate in the GTP bound state of the dimers. In contrast, the end of the tubule that is more likely to shrink is referred to as the minus-end. However, the *dynamic instability* models ignore all these details, instead treating the tubule as a polymer with subunits that are being added or removed at certain velocities (Fig. 8.2b). The power of this model however lies in its ability to precisely predict the length distributions of experimental microtubules, based on these simple parameters (Verde et al. 1992).



**Fig. 8.2** *The dynamics of microtubules.* (a) The dynamics of microtubules can be described at a molecular level with a model of dimer addition and removal. The GTP-cap model proposes that non-dynamic forms of the tubulin tube are GDP bound while those at the ends form a GTP-bound cap. (b) A more abstract treatment often used in Monte Carlo simulations of microtubule dynamics is of microtubules as a long polymer chain with subunits falling off and being added. The process of transition from growing to shrinking is called catastrophe and shrinking to growing rescue. The fibers transition in this model spontaneously at experimentally measured frequencies

Polymerizing microtubules can generate forces as has been seen in experiments with centrosomes placed in geometrically restricted chambers – the centrosomes reach the geometric centres of the chambers and oscillate (Holy et al. 1997). Experimental measurement has shown that the microtubule dynamics are affected by the force acting on them, i.e., if a microtubule comes into contact with an immovable obstacle, that microtubule will undergo shrinkage (Janson et al. 2003). A further study quantified the forces of microtubules in bundles by optical tweezer measurements. It demonstrated that force generation and single microtubule catastrophes could be coupled, leading to a model of microtubule force generation and oscillation of lengths of bundles (Laan et al. 2008).

Increasing length of microtubules appears to decrease rescue and increase catastrophe frequency of microtubules nucleated from centrosomes (Dogterom et al. 1996a). Such a mechanism has been found to be necessary to add to simulations of *S. pombe* microtubules in order to match the simulated organization with that experimentally measured (Foethke et al. 2009).

### 8.3 Microtubule-Motor Interactions

The spatial organization of microtubules in the spindle assembly process is primarily driven in vivo by energy dependent molecular motor proteins. The microtubules have an orientation of plus- and minus-ends. Similarly, the movement of motors is seen to fall into two classes of motors that move to either plus- or minus-ends. Typically these interactions have been studied in scales ranging from single motor-microtubule activities to the role of collective behaviour.

### 8.3.1 *Microtubule Gliding Assays*

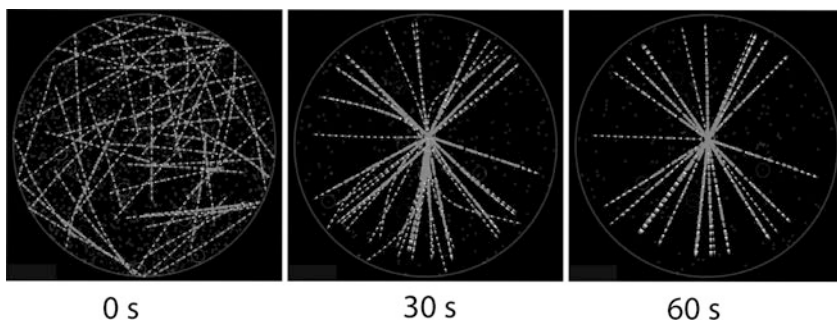
This assay is used typically to assess the kinetics of movement of microtubules by motors that are bound to a surface. A model of such “gliding assays” has been developed to understand the dependencies between the motors and microtubules that lead to the statistical properties of motion (Duke et al. 1995). Furthermore, defects in the motor density that lead to spiralling of the microtubule on the surface has been modeled to estimate the forces exerted by the motor proteins (Bourdieu et al. 1995). More recently, the collective behaviour of motors has been shown to depend on the rigidity of the motor in the case of kinesin where a shortened version of the motor decreases microtubule motility with increasing density of the motor (Bieling et al. 2008). Thus, loose coupling appears to be essential for cooperative activity of motor driven microtubule movement.

### 8.3.2 *Motor Mechanics*

Microtubule dependent motors are classified as either being plus- or minus-ended, depending on the direction of the microtubule they preferentially move towards. Most kinesins are plus-end directed, while most dyneins are minus-end directed. The step sizes and force produced by single motors has been estimated in the past using single molecule optical tweezer experiments for kinesin (Svoboda et al. 1993; Visscher et al. 1999). How these motor mechanics affect spindle assembly is more complex. One of the explanations is that a force-balance between inward and outward forces determines the length of the spindle as demonstrated by studies in *Drosophila* embryos by live imaging and modeling (Cytrynbaum et al. 2005). The inward forces are derived from minus-end directed motors (e.g., dynein) as demonstrated in the centering of microtubules in fish melanocyte fragments (Cytrynbaum et al. 2003). Experimental testing of these predictions demonstrated the need to include nuclear stretching in the generation of inward force (Cytrynbaum et al. 2005). A more minimal motor-microtubule modeling approach explored the steady state effects of motors on two centrosomal asters and demonstrated sliding forces could generate antiparallel overlaps and result in bipolar structures resembling spindles (Nedelec 2002). However, the hypothesized coupled dimeric plus- and minus-ended motor complexes are yet to be experimentally discovered.

### 8.3.3 *Microtubule-Motor Patterns*

Looking at collective behaviour, when mixtures of multi-headed kinesin were mixed with randomly nucleated microtubules in vitro, asters form in unconstrained and spatially restricted chambers by multiple pathways (Nedelec et al. 1997). This can also be simulated based on basic principles as seen in Fig. 8.3. Subsequently a model



**Fig. 8.3** *Spontaneous microtubule pattern formation.* Spontaneous patterns formed in a rigid circle in a 2D simulation of randomly distributed microtubules ( $n = 60$ ) are organized by plus-ended dimeric kinesin motors ( $n = 3,000$ ) into a radial centered array. The grey dots are motors which are either freely diffusing or can bind to and act on microtubules with a rate of attachment of  $10 \text{ s}^{-1}$  and detachment of  $0.5 \text{ s}^{-1}$ . (Simulations performed using Cytosim a C++ library described in Nedelec and Foethke (2007))

solution of a *convection–diffusion equation*<sup>1</sup> was developed, showing distribution of molecular motor densities on centrosomal asters can be described by continuously varying exponents (Nedelec et al. 2001). Performing stochastic simulations and experimental measurements on fluorescently tagged kinesin molecules in the presence of centrosomal asters both produced a similar outcome. When microtubules were mixed with oligomeric motors of opposite directionality – *ncd* (non claret disjunction, a minus-end kinesin motor protein) and kinesin-5 a plus-end kinesin – microtubules demonstrated spontaneous emergence of diverse patterns – spirals, whorls and asters. A theoretical model predicted these patterns, thus demonstrating that such patterns are the result of fundamental physical principles (Surrey et al. 2001). The role of microtubule-dynein interactions to produce an aster was explored using computer simulations applied to melanocyte granule aggregation (Cytrynbaum et al. 2004). A more general approach to phase transitions and ordering of microtubule filaments of motor systems has been taken by modeling cooperative behaviour of microtubule in motility assays using *Langevin dynamics* (Kraikivski et al. 2006). Such simulations use stochastic components to represent omitted variables, in order to enable abstract models. The typical Langevin equation of motion in such models as implemented in Cytosim (Nedelec and Foethke 2007) reads:

$$d\mathbf{x} = \mu F(\mathbf{x}, t)dt + dB(t), \quad (8.2)$$

where  $F(\mathbf{x}, t)$  represents the forces acting on the vector of points  $\mathbf{x}$  at time  $t$ ,  $B(t)$  are the random molecular collisions leading to Brownian motion and  $\mu$  contains the mobility coefficient parameters. Most recently, a model of plus- and minus-end motor mechanics was combined with a model of short microtubules that are

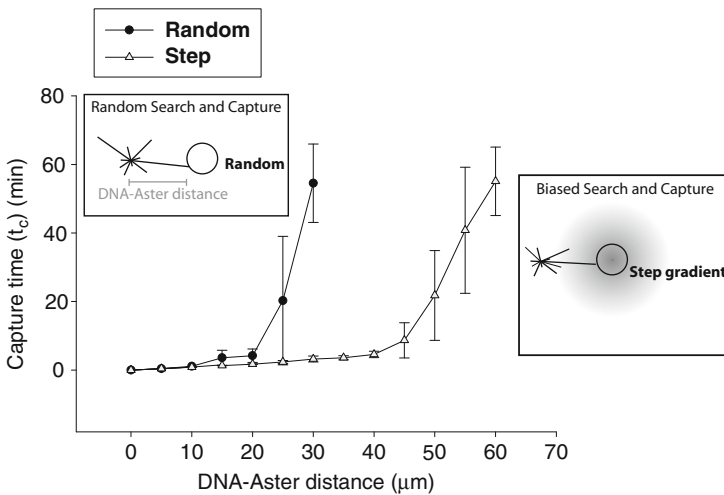
<sup>1</sup> Convection refers to directional movement of molecular motors, diffusion refers to random movement due to diffusion.

nucleated from chromosomes as well as on pre-existing microtubules. Such a system was shown to self-organize to a bipolar spindle structure in a model referred to as “slide and cluster” (Burbank et al. 2007). These approaches of modeling the patterns formed by microtubule motor systems appear to produce structures qualitatively similar to mitotic spindles and are represent a promising step towards a theory of spindle assembly.

## 8.4 Chromosome Dynamics

### 8.4.1 Search and Capture

During prometaphase, microtubules are more dynamic than in interphase, and are apparently efficient in finding chromosomes – a process termed as “search and capture”. The time from prometaphase to metaphase plate formation is typically < 30 min. The initial event of spindle assembly that involves both centrosomes and chromosomal DNA is thought to be result of randomly growing and shrinking microtubules (search), which might get immobilized on chromosome (capture) (Fig. 8.4 inset). This “search and capture” has to occur in a short time window with



**Fig. 8.4** *Search and capture*. Search and capture is the process by which a microtubules first encounters a chromosomal patch a fixed distance away – and the time taken is capture time ( $t_c$ ). A simulation of microtubules nucleated from a centrosomal aster perform random search and capture when the process purely depends on stochastic fluctuations in microtubule length (*left inset*), while in biased search and capture a gradient around the chromosomes preferentially increases microtubules growth locally (*right inset*).  $t_c$  in random search and capture remains biologically realistic (<10 min) even when the distance between aster and chromatin is <45  $\mu\text{m}$ , while  $t_c$  for random search and capture is already >10 min for distances of >25  $\mu\text{m}$

a high degree of accuracy since any mistake might result in loss of chromosomes or missegregation. Models that proposed the optimal values of dynamic instability parameters which would allow for reliable search and capture for small cells were proposed (Hill 1985; Holy and Leibler 1994). However, for cells larger than  $25\text{ }\mu\text{m}$  sizes the process of “random” search and capture appears to be insufficient (Fig. 8.4). These calculations are made considering the encounter of a single-microtubule with a single chromosomal patch of  $5\text{ }\mu\text{m}$  diameter. However, in a model that considers all 46 human chromosomes in a 3D environment in a typical cell,  $t_c$  is limited by the delay in those chromosomes capturing at least one microtubule that did not initially capture a microtubule. This delay increases the waiting time before anaphase to  $\sim 1\text{ h}$ . To overcome this, short range gradients of microtubule regulators (range  $\sim 5\text{ }\mu\text{m}$ ) have been proposed that can change the capture time to more realistic values (Wollman et al. 2008). These gradients are based on experimentally measured chromosomal gradients of RanGTP<sup>2</sup> (Kalab et al. 2002) and complexes (Kalab et al. 2006; Caudron et al. 2005). Most recently this author has shown with co-workers how a step-gradient of stabilization maintains capture times under 10 min for distances between chromosomal DNA and centrosome of up to  $45\text{ }\mu\text{m}$  (Athale et al. 2008) (Fig. 8.4). In starfish oocytes where distances between chromosomes and centrosomes can exceed  $45\text{ }\mu\text{m}$ , experimental work has demonstrated the role of a contractile, the acto-myosin network, that initially causes congression of chromosomes (Lenart et al. 2005). Thus such a simple model of “search and capture” has provided insights into a module of spindle assembly, that has contributed to our understanding of the more complex in vivo situation.

### 8.4.2 *Metaphase Plate Formation*

Chromosomes form the structure which is to be separated by the spindle apparatus. The process of attachment of each kinetochore of a pair of chromosomes such that the kinetochores lie equidistant from both poles at the metaphase plate in a spindle is referred to as congression. Mechanisms where chromosomes play either a passive or an active role in chromosome congression have been proposed. A physical model proposed that interactions of astral microtubules with the cell surface produces a balance of forces that leads to spindle placement at positions of stable equilibrium where forces with chromosomes simply reacting to this balance of forces (Bjerknes 1986). As an alternative, microtubule polymerization and depolymerization at the kinetochore was proposed to be transduced into chromosome movement (Hill 1985). A qualitative model of chromosome displacement in prometaphase proposed chromosomes move in two states: (a) energy driven active motion, and (b) neutral diffusion driven kinetochore movements, and a stochas-

---

<sup>2</sup> Ran = Ras-related nuclear protein.

tic switching between these two states (Khodjakov et al. 1999). Such a simple model provided good agreement with the data that quantified the time dynamics of chromosome movement. An improved mechanical model explicitly considered the polar ejection forces (forces due to astral microtubules) and balance of kinetochore microtubule forces from the spindle poles to generate the same behaviour without making assumptions about stochastic switching of states, relying instead on measured microtubule dynamics (Joglekar and Hunt 2002). When chromosomes congress at metaphase, the question arises if this compaction is random or follows any specific order. An experiment that used bleach marks in fluorescently labelled histone proteins to mark the orientation of chromatin DNA in live dividing cells was designed to test if the axes parallel and perpendicular to the metaphase plate axis (Fig. 8.1b) were maintained after anaphase, i.e., daughter nuclei formation. Only a model that assumes the maintenance of global spatial order of chromosomes (treated as rigid spheres) in the formation of the metaphase plate can explain why the bleach mark perpendicular to the metaphase plate is maintained from a prophase nucleus into its daughter nuclei (Gerlich et al. 2003). Models at various level of description are thus beginning to shed light on the genetically most important components of the spindle assembly process.

## 8.5 Reaction-Diffusion Gradients of Microtubule Dynamics Regulation

Very early on, experimental work by Bataillon (1912) showed that on injection of somatic nuclei into frog eggs they adopted the same mitotic state as the egg. Many workers elaborated on this to develop the concept of a *global cytoplasmic state*. However, in *Xenopus* egg extracts, mitotic spindles were shown to assemble in the absence of centrosomes and kinetochores (Heald et al. 1996), leading to the question whether the structures were all being affected by physical contact with chromatin or if chromatin was modifying the *local* state of cytoplasm. Through experimentation it could be shown that local modification of microtubules by gradients of protein phosphorylation was the most likely mechanism (Karsenti et al. 1984; Dogterom et al. 1996b). Some of the regulatory gradients determining spindle assembly have been modeled and are summarized here.

### 8.5.1 *Stathmin*

Initially, a known regulator of microtubule dynamics, stathmin which induces microtubule catastrophes, was shown on hyperphosphorylation to suppress catastrophe frequencies of microtubule dynamics (Andersen et al. 1997). Eventually, an elegant FRET sensor was developed that allowed the direct visualization spatial patterns of stathmin phosphorylation and it was indeed found to localize in a gradient

around chromosomes in spindle assembly (Niethammer et al. 2004). This gradient was modeled using an analytical solution of a the reaction-diffusion system of phosphorylation due to cytoplasmic phosphatases and a cell membrane bound kinase (Brown and Kholodenko 1999). The model predicted a gradient length of 4–8  $\mu\text{m}$ , corresponding well with the experimental extent of the gradient.

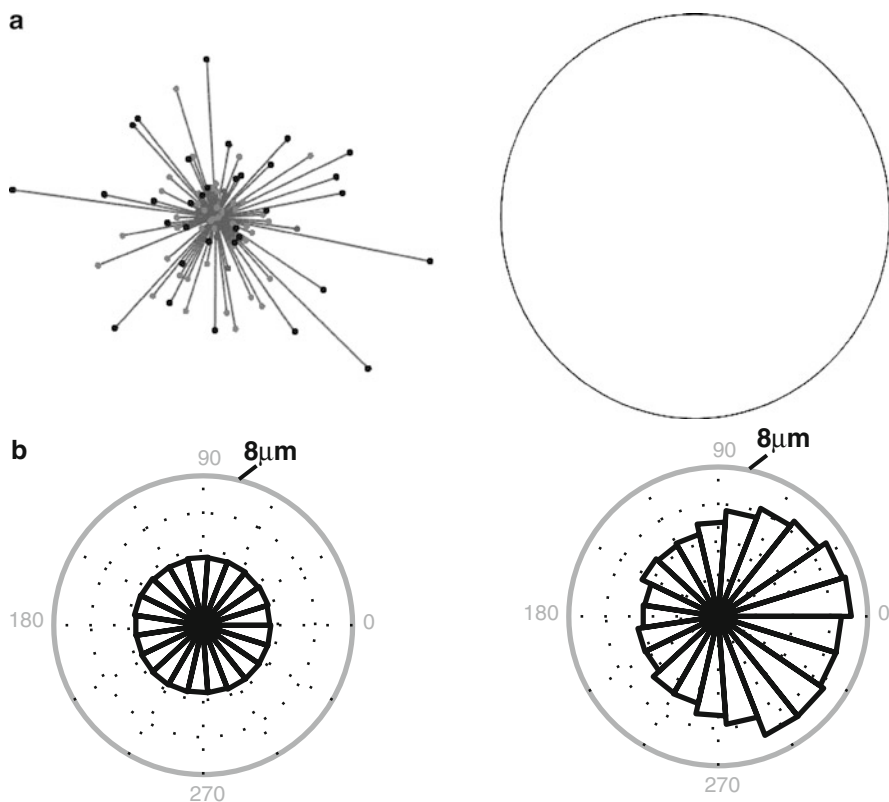
### 8.5.2 *RanGTP Nucleation and Stabilization Gradients*

Gradients of protein phosphorylation have been measured and modeled in the interphasic nucleo-cytoplasmic transport machinery where the nuclear localization signal (NLS) protein shuttling is governed by nuclear enriched RanGTP and cytoplasmic RanGDP (Fig. 8.6a) due to preferential localization of the kinase on chromatin, and the phosphatase in the bulk cytoplasm (Mattaj and Englmeier 1998; Gorlich and Kutay 1999; Gorlich et al. 2003). The cargo protein to be targeted to the nucleus is bound non-covalently to a class of carrier proteins, the importins. The release of the cargo protein from importins depends on binding with Ran in its GTP bound form (RanGTP).

In mitosis, in the absence of a nuclear membrane, the RanGTP system with kinase on chromatin and the phosphatase RanGAP (Ran GTPase activating protein) in the cytoplasmic bulk produces a short range gradient of phosphorylation which could be measured (Kalab et al. 2006). The induction of nucleation and stabilization of microtubules by chromatin requires RanGTP. One of the nucleation factors has been found to be RanGTP dependent TPX2 (TPX2 = targeting protein for XKLP2) (Gruss et al. 2001, 2002) that is thought to be released in a short-range gradient around chromosomes. The range of the RanGTP-Importin complex was measured to estimate the range of such Importin-released factors and a value of approximately 7  $\mu\text{m}$  was arrived at (Caudron et al. 2005; Kalab et al. 2006). This gradient range is also in agreement with calculations using the reaction rates and assumed diffusion coefficients in the analytical solution for sub-cellular phosphorylation gradients (Brown and Kholodenko 1999) (Fig. 8.6b).

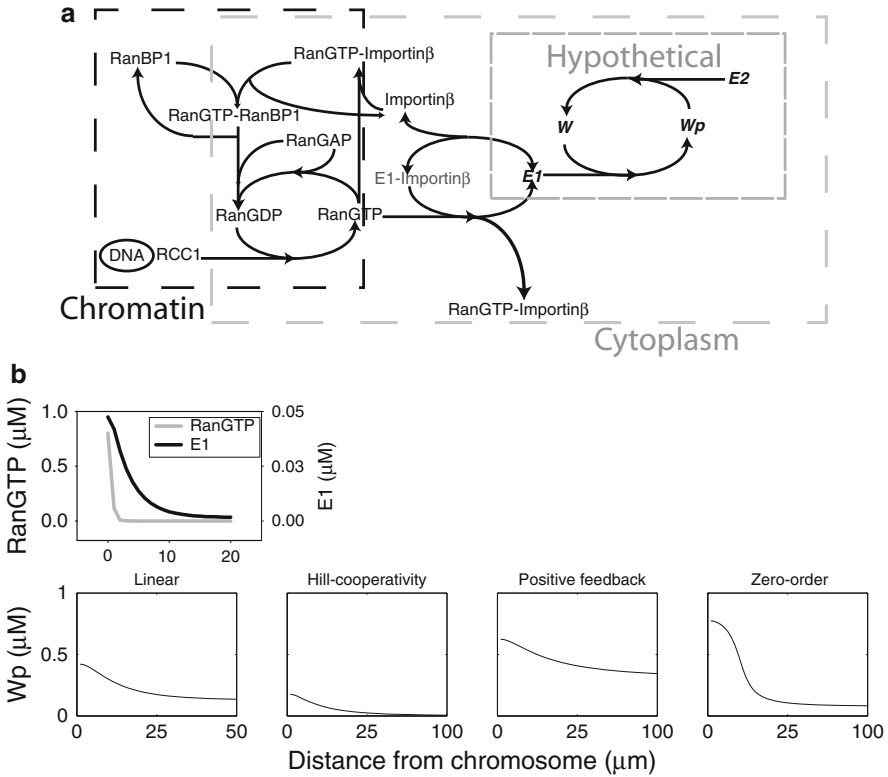
### 8.5.3 *Long-Range Stabilization Gradients*

The measured gradients of RanGTP and its complex with Importin were measured to be relatively short-range (5–10  $\mu\text{m}$ ), as also expected from calculations (Fig. 8.6). However, experiments had shown that centrosomal microtubules display polarized growth due to long-range stabilization, originating from chromosomes in a RanGTP dependent manner. This gradient has a longer range, in the order of 20  $\mu\text{m}$  (Carazo-Salas and Karsenti 2003). In order to explain this discrepancy a model was developed by the author (Athale et al. 2008) of the dynamic instability of centrosomal microtubules using a stochastic simulation of finite numbers of microtubules



**Fig. 8.5** *Simulation of a radial aster steady state mean distribution of astral microtubules. (a)* The simulation of a radial aster in the presence of a circular chromatin structure. The *dark dots* at the microtubule ends indicate those microtubule plus-ends that are growing, and the *light-grey dots* indicate those that are shrinking. **(b)** The steady state mean distribution of astral microtubules is symmetric for the case without a stabilization gradient (*left*) and asymmetric in the presence of a gradient (*right*)

growing under the influence of a gradient of stabilization (Fig. 8.5). The effect of the gradient was based on experimental values of RanGTP modification of microtubule dynamics (Table 8.1). Its shape was derived from assumptions about the gradient forming reactions. By comparing the simulation of aster asymmetry with published and fresh experimental data, we were surprised to find that expected models of phosphorylation–dephosphorylation gradients (Brown and Kholodenko 1999) did not work, and a long range step-like gradient needed to be invoked. How such a step-like gradient shape could be generated in bulk cytoplasm however remained to be answered. We approached the problem by adding a hypothetical reaction module to the existing RanGTP reaction-diffusion network and solving the system as a partial differential equation (PDE) in one spatial dimension, assuming a spherical geometry around chromatin. This module included a substrate  $W$  which is initially



**Fig. 8.6** Reaction-network of the chromatin mediated components and radial gradient from the chromatin surface into the cytoplasm. (a) The reaction-network of the chromatin mediated components that generate a RanGTP protein gradient around chromosomes. The addition of a hypothetical reaction network was simulated to explain experimental data. (b) The radial gradient from the chromatin surface into the cytoplasm produced by RanGTP and E1 are short range, while the hypothetical phosphorylated substrate Wp can form a long-range gradient for some reaction topologies (adapted from the author's work in (Athale et al. 2008))

unphosphorylated but can be phosphorylated by a kinase E1 to Wp (Fig. 8.6a). The phosphorylated form can in turn be dephosphorylated by a phosphatase E2, forming a cyclic reaction network. The E1 was assumed to be sequestered by importins and released in the presence of RanGTP, just like other NLS factors (parameters used in Table 8.1). Testing four different kinetic regimes, linear, Hill-cooperativity, positive-feedback and zero-order ultrasensitivity, we could show that the zero-order ultrasensitive network was most capable of producing the effects that agreed qualitatively with the experimental data of microtubule asymmetry near chromatin (Fig. 8.6b). Recently a RanGTP dependent factor that stabilizes microtubules has been identified to be a kinase CDK11 (Yokoyama et al. 2008). It is quite likely that this might be the identity of the hypothesized E1 kinase, demonstrating the power of predictive modeling to design experiments and gain insights.

**Table 8.1** Parameters for the stabilization of microtubules by the RanGTP gradient generating system

| Microtubule dynamic instability parameters                                 |                        |                                            |                            |
|----------------------------------------------------------------------------|------------------------|--------------------------------------------|----------------------------|
| $f_{cat}$                                                                  | Mitotic                | 0.0498/s                                   | — <sup>a</sup>             |
|                                                                            | Chromosome stabilized  | 0.03 s <sup>−1</sup>                       | — <sup>a</sup>             |
| $f_{res}$                                                                  | Mitotic                | 0.0048/s                                   | — <sup>a</sup>             |
|                                                                            | Chromosome stabilized  | 0.012 s <sup>−1</sup>                      | — <sup>a</sup>             |
| $v_g$                                                                      | Mitotic and stabilized | 0.196 μm/s                                 | — <sup>a</sup>             |
| $v_s$                                                                      | Mitotic and stabilized | 0.325 μm/s                                 | — <sup>a</sup>             |
| Reaction diffusion parameters of hypothesized reaction system <sup>b</sup> |                        |                                            |                            |
| Species                                                                    | Localization           | Diffusion Coefficient (μm <sup>2</sup> /s) | Initial concentration (μM) |
| E1 (Kinase)                                                                | Cytoplasmic            | 10                                         | 1                          |
| E2 (Phosphatase)                                                           | Cytoplasmic            | 10                                         | 0.1                        |
| Wt (Total substrate)                                                       | Cytoplasmic            | 10                                         | 1                          |
| Wp (Phosphorylated substrate)                                              | Cytoplasmic            | 10                                         | 0                          |
| WE1 (Substrate-kinase complex)                                             | Cytoplasmic            | 5                                          | 0                          |
| WpE2 (Product-phosphatase complex)                                         | Cytoplasmic            | 5                                          | 0                          |

<sup>a</sup>Wilde et al. (2001), Carazo-Salas et al. (2001)

<sup>b</sup>Athale et al. (2008)

8.6 Outlook

The preceding sections have summarized how a problem of cellular pattern formation, namely the assembly of a spindle machinery has been tackled using methods ranging from microscopy, molecular biophysics, surface chemistry, partial differential equations, monte-carlo simulations and molecular biology. All the findings have been built gradually into theories of modules of the system. Often older theories have also been modified by new experimental findings and experimental techniques, as in the example of the “random search and capture” model being replaced by a “biased search and capture” model in spindle assembly. As is often true of models, the simplifications however ignore the obvious details such as search and capture not being the only mechanism necessary for spindle assembly, as demonstrated in experiments where spindles assemble in the absence of centrosomes (Heald et al. 1996).

Thus, we find also new principles of pattern formation being discovered during this process of iterative modeling and experimental comparison. An example is the case of measurements of aster anisotropy in experiments (Carazo-Salas and Karsenti 2003; Dogterom et al. 1996a) which suggested a RanGTP dependent chromosomal stabilization gradient. It was initially assumed the RanGTP dependent release of a factor would be sufficient to produce this effect of stabilization based on measurements and models of the gradient forming reactions (Caudron et al. 2005; Wollman et al. 2005). However, it was only after a hybrid simulation of the reaction-diffusion dynamics and the stochastic microtubule growth dynamics were modeled that it became apparent that this model was not sufficient to explain the experimental data,

leading to both new experimental design to confirm this, as well as a new model to propose mechanisms that might fulfil the criteria (Athale et al. 2008).

Taken together, the spindle assembly process represents a paradigm for cellular pattern formation, since it involves all the common processes that influence cellular patterns – chemical reaction-diffusion, cytoskeletal mechanics, motor driven motility, genetic regulation and self-organization. Additionally many of the molecular actors have been identified by a combination of genetics, biochemistry and reconstitution of pure proteins to mimic the in vivo condition. The question still remains whether an overarching model can be found to describe the assembly and functioning of this complex molecular machine.

In future such an approach of simulating modules of the spindle assembly system will probably intensify driven by better and more quantitative experimental techniques. This will not just make the models more valid and better tested, but also open the way to modeling that involves combinations of different modules. Eventually it can be envisaged that the ambitious goal of modeling the whole structure might become more realistic. At that point predictions could then become multi-scale, stretching from molecular interaction dynamics to functional properties. The integration of modeling and experimentation using quantitative biophysical tools will thus pave the way to a systems-level understanding of not just spindle assembly, but possibly other cellular processes that use similar modules such as cell polarization, migration and differentiation.

**Acknowledgements** C.A.A. was funded by BioMS and hosted in the lab of Eric Karsenti at EMBL Heidelberg. The author is grateful to Eric Karsenti for discussions and Francois Nedelec for providing access to the C++ tool Cytosim for simulations of microtubule and motor dynamics.

## References

- A. Akhmanova and M. O. Steinmetz. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nat Rev Mol Cell Biol*, 9:309–322, 2008
- S. S. Andersen, A. J. Ashford, R. Tournebise, O. Gavet, A. Sobel, A. A. Hyman, and E. Karsenti. Mitotic chromatin regulates phosphorylation of Stathmin/Op18. *Nature*, 389:640–643, 1997
- C. A. Athale, A. Dinarina, M. Mora-Coral, C. Pugieux, F. Nedelec, and E. Karsenti. Regulation of microtubule dynamics by reaction cascades around chromosomes. *Science*, 322: 1243–1247, 2008
- E. Bataillon. La parthenogenese des Amphibiens et la fcondation chimique de Loeb. *Ann Sci Nat Zool*, 16:249–307, 1912
- P. Bieling, L. Laan, H. Schek, E. L. Munteanu, L. Sandblad, M. Dogterom, D. Brunner, and T. Surrey. Reconstitution of a microtubule plus-end tracking system in vitro. *Nature*, 450:1100–1105, 2007
- P. Bieling, I. A. Telley, J. Piehler, and T. Surrey. Processive kinesins require loose mechanical coupling for efficient collective motility. *EMBO Rep*, 9:1121–1127, 2008
- M. Bjerknes. Physical theory of the orientation of astral mitotic spindles. *Science*, 234:1413–1416, 1986
- L. Bourdieu, T. Duke, M. B. Elowitz, D. A. Winkelmann, S. Leibler, and A. Libchaber. Spiral defects in motility assays: A measure of motor protein force. *Phys Rev Lett*, 75:176–179, 1995

- G. C. Brown and B. N. Kholodenko. Spatial gradients of cellular phospho-proteins. *FEBS Lett*, 457:452–454, 1999
- K. S. Burbank, T. J. Mitchison, and D. S. Fisher. Slide-and-cluster models for spindle assembly. *Curr Biol*, 17:1373–1383, 2007
- R. E. Carazo-Salas and E. Karsenti. Long-range communication between chromatin and microtubules in *Xenopus* egg extracts. *Curr Biol*, 13:1728–1733, 2003
- R. E. Carazo-Salas, O. J. Gruss, I. W. Mattaj, and E. Karsenti. Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. *Nat Cell Biol*, 3:228–234, 2001
- M. Caudron, G. Bunt, P. Bastiaens, and E. Karsenti. Spatial coordination of spindle assembly by chromosome-mediated signaling gradients. *Science*, 309:1373–1376, 2005
- Y. Chen and T. L. Hill. Theoretical treatment of microtubules disappearing in solution. *Proc Natl Acad Sci USA*, 82:4127–4131, 1985
- T. Clausen and K. Ribbeck. Self-organization of anastral spindles by synergy of dynamic instability, autocatalytic microtubule production, and a spatial signaling gradient. *PLoS ONE*, 2:e244, 2007
- E. N. Cytrynbaum, J. M. Scholey, and A. Mogilner. A force balance model of early spindle pole separation in *Drosophila* embryos. *Biophys J*, 84:757–769, 2003
- E. N. Cytrynbaum, V. Rodionov, and A. Mogilner. Computational model of dynein-dependent self-organization of microtubule asters. *J Cell Sci*, 117:1381–1397, 2004
- E. N. Cytrynbaum, P. Sommi, I. Brust-Mascher, J. M. Scholey, and A. Mogilner. Early spindle assembly in *Drosophila* embryos: role of a force balance involving cytoskeletal dynamics and nuclear mechanics. *Mol Biol Cell*, 16:4967–4981, 2005
- M. Dogterom, A. C. Maggs, and S. Leibler. Diffusion and formation of microtubule asters: physical processes versus biochemical regulation. *Proc Natl Acad Sci USA*, 92:6683–6688, 1995
- M. Dogterom, M. A. Félix, C. C. Guet, and S. Leibler. Influence of M-phase chromatin on the anisotropy of microtubule asters. *J Cell Biol*, 133:125–140, 1996a
- M. Dogterom, M. A. Félix, C. C. Guet, and S. Leibler. Influence of M-phase chromatin on the anisotropy of microtubule asters. *J Cell Biol*, 133:125–140, 1996b
- T. Duke, T. E. Holy, and S. Leibler. “Gliding assays” for motor proteins: A theoretical analysis. *Phys Rev Lett*, 74:330–333, 1995
- D. W. Ehrhardt. Straighten up and fly right: microtubule dynamics and organization of non-centrosomal arrays in higher plants. *Curr Opin Cell Biol*, 20:107–116, 2008
- W. Flemming. *Zellsubstanz, Kern und Zellteilung*. Verlag Vogel, Leipzig, 1882
- D. Foethke, T. Makushok, D. Brunner, and F. Ndlec. Force- and length-dependent catastrophe activities explain interphase microtubule organization in fission yeast. *Mol Syst Biol*, 5:241, 2009
- D. Gerlich, J. Beaudouin, B. Kalbfuss, N. Daigle, R. Eils, and J. Ellenberg. Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell*, 112:751–764, 2003
- D. Gorlich and U. Kutay. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol*, 15:607–660, 1999
- D. Gorlich, M. J. Seewald, and K. Ribbeck. Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. *EMBO J*, 22:1088–1100, 2003
- O. J. Gruss, R. E. Carazo-Salas, C. A. Schatz, G. Guarguaglini, J. Kast, M. Wilm, N. Le Bot, I. Vernos, E. Karsenti, and I. W. Mattaj. Ran induces spindle assembly by reversing the inhibitory effect of importin alpha on TPX2 activity. *Cell*, 104:83–93, 2001
- O. J. Gruss, M. Wittmann, H. Yokoyama, R. Pepperkok, T. Kufer, H. Sillj, E. Karsenti, I. W. Mattaj, and I. Vernos. Chromosome-induced microtubule assembly mediated by TPX2 is required for spindle formation in HeLa cells. *Nat Cell Biol*, 4:871–879, 2002
- R. Heald, R. Tournebise, T. Blank, R. Sandaltzopoulos, P. Becker, A. Hyman, and E. Karsenti. Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature*, 382:420–425, 1996

- T. L. Hill. Introductory analysis of the GTP-cap phase-change kinetics at the end of a microtubule. *Proc Natl Acad Sci USA*, 81:6728–6732, 1984
- T. L. Hill. Theoretical problems related to the attachment of microtubules to kinetochores. *Proc Natl Acad Sci USA*, 82:4404–4408, 1985
- T. E. Holy and S. Leibler. Dynamic instability of microtubules as an efficient way to search in space. *Proc Natl Acad Sci USA*, 91:5682–5685, 1994
- T. E. Holy, M. Dogterom, B. Yurke, and S. Leibler. Assembly and positioning of microtubule asters in microfabricated chambers. *Proc Natl Acad Sci USA*, 94:6228–6231, 1997
- M. E. Janson, M. E. de Dood, and M. Dogterom. Dynamic instability of microtubules is regulated by force. *J Cell Biol*, 161:1029–1034, 2003
- A. P. Joglekar and A. J. Hunt. A simple, mechanistic model for directional instability during mitotic chromosome movements. *Biophys J*, 83:42–58, 2002
- P. Kalab, K. Weis, and R. Heald. Visualization of a Ran-GTP gradient in interphase and mitotic *Xenopus* egg extracts. *Science*, 295:2452–2456, 2002
- P. Kalab, A. Pralle, E. Y. Isacoff, R. Heald, and K. Weis. Analysis of a RanGTP-regulated gradient in mitotic somatic cells. *Nature*, 440:697–701, 2006
- E. Karsenti, J. Newport, R. Hubble, and M. Kirschner. Interconversion of metaphase and interphase microtubule arrays, as studied by the injection of centrosomes and nuclei into *Xenopus* eggs. *J Cell Biol*, 98:1730–1745, 1984
- A. Khodjakov, I. S. Gabashvili, and C. L. Rieder. “Dumb” versus “smart” kinetochore models for chromosome congression during mitosis in vertebrate somatic cells. *Cell Motil Cytoskeleton*, 43:179–185, 1999
- P. Kraikivski, R. Lipowsky, and J. Kierfeld. Enhanced ordering of interacting filaments by molecular motors. *Phys Rev Lett*, 96:258103, 2006
- L. Laan, J. Husson, E. L. Munteanu, J. W. Kerssemakers, and M. Dogterom. Force-generation and dynamic instability of microtubule bundles. *Proc Natl Acad Sci USA*, 105:8920–8925, 2008
- P. Lenart, C. P. Bacher, N. Daigle, A. R. Hand, R. Eils, M. Terasaki, and J. Ellenberg. A contractile nuclear actin network drives chromosome congression in oocytes. *Nature*, 436:812–818, 2005
- I. W. Mattaj and L. Englmeier. Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem*, 67:265–306, 1998
- H. Meinhardt and P. A. J. de Boer. Pattern formation in *Escherichia coli*: A model for the pole-to-pole oscillations of Min proteins and the localization of the division site. *Proc Natl Acad Sci USA*, 98:14202–14207, 2001
- T. Mitchison and M. Kirschner. Dynamic instability of microtubule growth. *Nature*, 312:237–242, 1984
- F. Nedelec. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. *J Cell Biol*, 158:1005–1015, 2002
- F. Nedelec, T. Surrey, and A. C. Maggs. Dynamic concentration of motors in microtubule arrays. *Phys Rev Lett*, 86:3192–3195, 2001
- F. J. Nedelec, T. Surrey, A. C. Maggs, and S. Leibler. Self-organization of microtubules and motors. *Nature*, 389:305–308, 1997
- Francois Nedelec and Dietrich Foethke. Collective langevin dynamics of flexible cytoskeletal fibers. *New Journal of Physics*, 9(11):427, 2007. URL <http://stacks.iop.org/1367-2630/9/427>
- P. Niethammer, P. Bastiaens, and E. Karsenti. Stathmin-tubulin interaction gradients in motile and mitotic cells. *Science*, 303:1862–1866, 2004
- P. Niethammer, I. Kronja, S. Kandels-Lewis, S. Rybina, P. Bastiaens, and E. Karsenti. Discrete states of a protein interaction network govern interphase and mitotic microtubule dynamics. *PLoS Biol*, 5:e29, 2007
- B. Raynaud-Messina and A. Merdes. Gamma-tubulin complexes and microtubule organization. *Curr Opin Cell Biol*, 19:24–30, 2007
- T. Surrey, F. Nedelec, S. Leibler, and E. Karsenti. Physical properties determining self-organization of motors and microtubules. *Science*, 292(5519):1167–71, 2001
- K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365:721–727, 1993

- F. Verde, M. Dogterom, E. Stelzer, E. Karsenti, and S. Leibler. Control of microtubule dynamics and length by cyclin a- and cyclin b-dependent kinases in xenopus egg extracts. *J Cell Biol*, 118(5):1097–108, 1992
- K. Visscher, M. J. Schnitzer, and S. M. Block. Single kinesin molecules studied with a molecular force clamp. *Nature*, 400:184–189, 1999
- C. E. Walczak, I. Vernos, T. J. Mitchison, E. Karsenti, and R. Heald. A model for the proposed roles of different microtubule-based motor proteins in establishing spindle bipolarity. *Curr Biol*, 8: 903–913, 1998
- A. Wilde, S. B. Lizarraga, L. Zhang, C. Wiese, N. R. Gliksman, C. E. Walczak, and Y. Zheng. Ran stimulates spindle assembly by altering microtubule dynamics and the balance of motor activities. *Nat Cell Biol*, 3:221–227, 2001
- R. Wollman, E. N. Cytrynbaum, J. T. Jones, T. Meyer, J. M. Scholey, and A. Mogilner. Efficient chromosome capture requires a bias in the 'search-and-capture' process during mitotic-spindle assembly. *Curr Biol*, 15:828–832, 2005
- R. Wollman, G. Civelekoglu-Scholey, J. M. Scholey, and A. Mogilner. Reverse engineering of force integration during mitosis in the *Drosophila* embryo. *Mol Syst Biol*, 4:195, 2008
- H. Yokoyama, O. J. Gruss, S. Rybina, M. Caudron, M. Schelder, M. Wilm, I. W. Mattaj, and E. Karsenti. Cdk11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. *J Cell Biol*, 180:867–875, 2008

# Chapter 9

## Cell-Centred Modeling of Tissue Behaviour

Rod Smallwood

### 9.1 Introduction: Towards a Virtual Cell Biology

The nature of the problem to be solved determines the modeling paradigm to use. The problem of interest is the development of normal structure and function, and the mechanisms that control homeostasis, in epithelial tissues, and the behaviour when the tissue is damaged (wound healing) or the normal homeostatic mechanisms are deranged (development of cancer). These processes are the result of the interaction of individual cells whose behaviour depends on internal and external information. The internal information store is the genetic material, and the external information is chemical and physical signals. This is clearly a cell-centred description of tissue behaviour, and equally clearly, an individual-based modeling paradigm, in which the cell is the individual, is appropriate. Which particular individual-based model to choose is less obvious, but in this case serendipity and logic led to the choice of Eilenberg's X-machine, which has sufficient power to enable multi-scale and multi-paradigm modeling. Modeling at the cell level has in general concentrated on either the biochemical aspects or the physical aspects, and not on the combination of the two, which is essential for understanding cellular behaviour. The choice of epithelial tissues was largely pragmatic – two excellent in vitro models of tissue were available, which was grown for human implantation (urothelial monolayers and skin); and epithelial tissues are about as simple a structure as one can find in biology – a very small number of cell types, no connective tissue, no nerve endings or blood vessels – and they have important barrier functions and are the source of all carcinomas. The long-term aim is to use the epithelial cell model as a starting point for a virtual stem cell, and to enable the expression of biological problems within a virtual cell biology which cell biologists can use for experimental work in parallel with wet biology. Ultimately, I think it is essential that the interface through the computer scientist or the physical scientist to the virtual cell biology is removed – the cell bi-

---

R. Smallwood (✉)  
Department of Computer Science, University of Sheffield, Regent Court,  
211 Portobello, Sheffield S1 4DP, UK  
e-mail: [r.smallwood@shef.ac.uk](mailto:r.smallwood@shef.ac.uk)

ologist should be able to perform and interpret virtual experiments just as they now perform and interpret wet biology experiments. This will require a sophisticated interface, the ability to build cells with appropriate functions in a modular manner and the ability to specify experimental conditions that are translated into starting and boundary conditions for the virtual biology. The aim is not to simulate biology, but to make testable predictions outside the parameter space of the input data. The reader is referred to [Smallwood \(2009\)](#) for a general review of epithelial tissue modeling, to [Walker et al. \(2004a, b, 2006a, b\)](#) and to [Sun et al. \(2007, 2008\)](#) for detailed descriptions of the models.

## 9.2 Can Computation Cope with Cellular Complexity?

The first problem that becomes apparent when discussing a cell-based model of tissue which incorporates molecular detail is the complexity – the probability space provided by  $\sim 30,000$  genes and  $10^5$  proteins in the  $10^{13}$  cells in the human body, combined with the range of spatial and temporal scales – from  $10^{-12}$ m (atom) to  $10^0$ m (organism), and  $10^{-9}$ s (molecular interaction) to  $10^9$ s (lifetime). [Feytmans et al. \(2005\)](#) have shown that, if 100 genes are required to code for each function, there are  $\sim 10^{289}$  possible combinations of 30,000 genes, and adding one more gene adds  $10^{287}$  new functions (there are about  $10^{80}$  atoms in the universe). Clearly some constraints have to be applied, and the constraints are evolutionary – biology is actually highly constrained and conservative (which is why *Arabidopsis* and *C. elegans* are useful model organisms). Despite the uncountable size of the combinatorial space, there are only about 200 different types of cells in mammals, and only 9 body plans in the Animalia. Similarly, despite the enormous numbers of proteins involved in signalling, there are only a handful of generic types of signalling – transport through ion channels and gap junctions, receptor–ligand binding, mechanotransduction, etc. This suggests that one strategy for reducing the apparent complexity would be to commence at a functional level, and introduce component (mechanistic) detail only where required for the problem in hand – an integrative, rather than reductionist, approach. Multiscale modeling can be used to introduce details only where required (and conversely, can be used to abstract away detail where it is not required). This also implies the use of the largest length and time scales which still give a reasonable representation of the underlying reality – an example is the use of large heterogeneous finite elements to provide mechanical details of the ventricles in cardiac models, in which only about 100 elements are required, but have properties that have been determined from sub-micrometer structural models. It also appears to be possible, with cellular-level models, to get representative behaviour from significantly fewer cells than are present in the biological model. A skin wound containing  $O(10^6)$  epithelial cells will not heal without grafting, but similar growth behaviour can be seen with  $O(10^4)$  cells in a computational model. This implies that the length scale at which cellular-level behaviour can be abstracted away to continuum-based tissue level behaviour may be as small as 1 mm.

### 9.2.1 *Being Generic: Function Versus Detail*

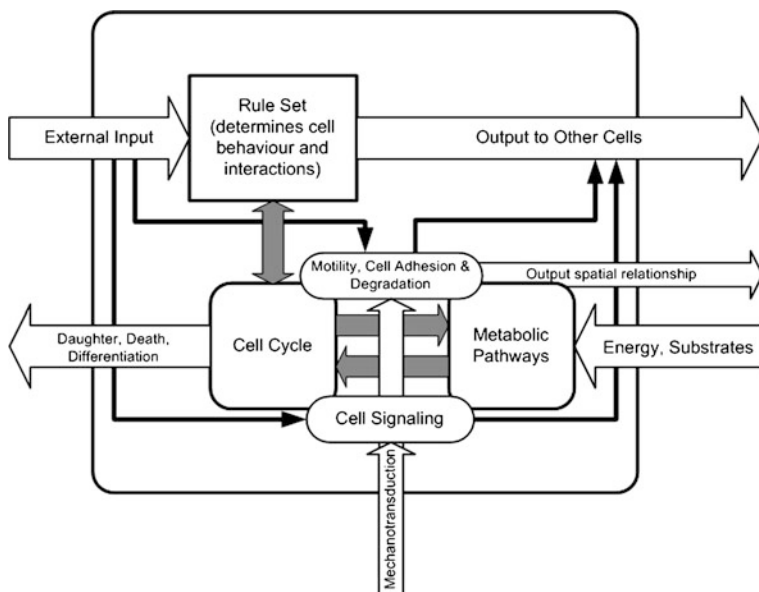
The signal transduction knowledge environment (see footnote 4) lists  $\sim 2,000$  molecules involved in signal transduction, and  $>100$  different pathways. At a generic level, these could be classified in a small number of groups:

- Movement of ions between adjacent cells via gap junctions.
- Binding of membrane-inserted proteins on the same cell or neighbouring cells.
- Binding of ligands diffusing through the extracellular space with membrane-inserted receptors.
- Mechanotransduction: mechanical to chemical transduction at focal adhesions; stretch-sensitive ion channels.
- ...

The description of these signalling processes does not require details of the proteins involved and their relationships, *unless* this is needed to explore how, for instance, the availability of a particular protein in the signalling chain (as a result of, e.g. sequestering of the protein in a store as a result of some other reaction which is not part of the signalling chain) affects the system level performance. Section 9.8.1 describes a model of a signalling chain which includes sequestration of one of the proteins, and Sect. 9.8.2 describes diffusion of ligands with receptor binding and trafficking.

## 9.3 Cells and Computation

The use of an individual-based approach leads to a 1:1 mapping between the biological cell and the cell's computational representation as a finite state machine (which I will refer to as an agent). Each agent has a local frame of reference – the individual agent behaviour is the result of internal events (cell cycle, growth, division, etc.) which are mediated by the external environment (physical interaction with other agents and the substrate; mechanical and chemical signalling). The agents have a physical location defined by the centroid of the cell on a continuous scale; a size, shape and mechanical properties; and exert and respond to forces resulting from growth and the formation of cell–cell and cell–substrate bonds. The physical behaviour could be embedded within the function set of the agent, but we have chosen to use a physical model which is separate from, but exchanges data with, the agent model. Extracellular signalling is of two generic types – communication with adjacent cells (gap junctions, binding of membrane-inserted ligands and receptors) which corresponds to local message passing between agents; and diffusible agents such as growth factors which provide a non-local signalling mechanism and require a diffusion model. Activities such as receptor trafficking need additional functions. Internal signalling, gene activation, etc. can be handled by individual or differential equation-based approaches as appropriate. The physical and chemical domains are linked by mechano-transduction, in which internal signalling is mediated by



**Fig. 9.1** Internal and external relationships of a cell/agent

externally applied forces. Although this chapter is primarily about epithelial cells, in a more general case, the behaviour could include electrical excitability and active length and shape change, electrical to mechanical coupling, and also mechano-electrical coupling (stretch-sensitive ion channels). This clearly requires complex, multi-scale models. The cellular functions are represented in Fig. 9.1, which is the basis for developing the set of states and functions that comprise the X-machine.

## 9.4 Developing a Multi-Scale Model

In principle, an individual-based model can be used at any level. The logic at the cellular level is that the transition from cellular behaviour to tissue behaviour is an emergent property of cellular interaction; so the modeling paradigm has to be able to reproduce this behaviour. At a higher level than a few tens of thousands of cells, one can argue that an ensemble average of the emergent properties results in a quasi-deterministic (or at least stochastic) process that one can more efficiently model using differential equations or continuum methods. This argument holds equally at lower levels than the cell – if more than a few thousand individual molecules are involved, the emergent detail can be abstracted away to an equation. There is an implicit assumption in so dealing with subcellular chemical events, which is that we are dealing with a well-stirred solution. This is clearly not the case within the cell, which is full of mechanisms for the local transport and sequestering of

molecules, but pragmatism and computational load have so far dictated that the local inhomogeneity is generally ignored. As the cellular mechanisms are evolved mechanisms, they presumably occupy at least a local optimum in a fitness landscape, and therefore their particular form is important; but I am not aware of any comprehensive attempt to relate subcellular morphology to function. In practice, it is clearly impossible for any one research group to build de novo a model of any reasonably sized biological system, and existing models and modeling paradigms will have to be used. The linking of different modeling paradigms will be discussed later.

## 9.5 The Agent Basis: The Communicating-Stream X-Machine

Individual-based models have been little used in engineering contexts (the major use has been in ecology – [Grimm \(1999\)](#) provides a critical review of individual-based modeling in ecology); so a familiarity with them cannot be assumed in the way it can be for differential equation or finite element-based models. I have elsewhere reviewed computational models of epithelial tissue ([Smallwood 2009](#)), so will limit my remarks to the choice of the X-machine as the formalism for the software agents. The requirements are as follows:

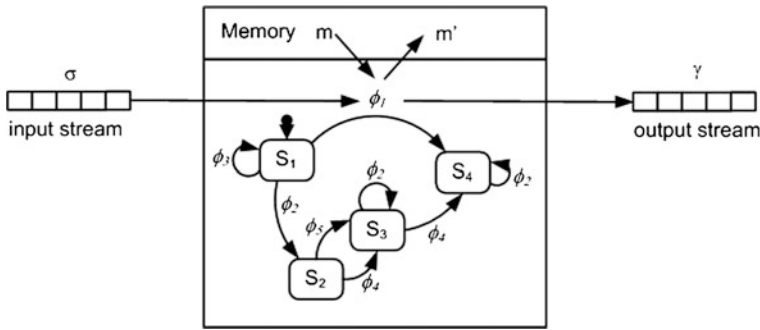
- The modeling paradigm has to be robust, as the goal is to develop computational models which are able to predict the effect of intervention in clinical problems.
- A 1:1 mapping between cells and agents was considered desirable; so the agent had to be capable of handling cellular complexity.
- Building models de novo of the whole of cell and tissue biology is clearly not feasible, so that linking or importing models of specific processes which had been built using different modeling paradigms was essential.
- The physical environment is an important determinant of cellular function; so linking or incorporating physical solvers was essential.
- The problems of state explosion had to be avoided.
- The number of cells in the smallest diameter non-healing skin wound is  $O(10^6)$ ; so it had to be possible to model this number of cells, which implies that parallelisation was essential.

The only candidate which met all the requirements was the X-machine, which was introduced by [Eilenberg \(1974\)](#). A good introduction and bibliography are provided by [Stannett \(2005\)](#). The communicating-stream X-machine (Fig. 9.2) is an extension of the basic deterministic stream X-machine ([Kefalas et al. 2003](#)), which is formally defined as an 8-tuple:

$$M = (\Sigma, \Gamma, Q, M, \Phi, F, q_0, m_0)$$

where

- $\Sigma$  and  $\Gamma$  are the input and output finite alphabets, respectively.
- $Q$  is the finite set of states.



**Fig. 9.2** The communicating X-machine [after Kefalas et al. (2003)].  $S_i$  are the states,  $\phi_i$  the functions operating on inputs  $\sigma$  and memory  $m$

- $M$  is the (possibly) infinite set called memory.
- $\Phi$  is the type of the machine  $M$ , a finite set of partial functions  $\phi$  that map an input and a memory state to an output and a new memory state  $\phi : \Sigma \times M \rightarrow \Gamma \times M$ .
- $F$  is the next state partial function that, given a state and a function from the type  $\Phi$ , denotes the next state.  $F$  is often described as a transition state diagram.  $F : Q \times \Phi \rightarrow Q$ .
- $q_0$  and  $m_0$  are the initial state and memory, respectively.

The addition of communication adds an input stream  $\sigma$  and an output stream  $\gamma$ . The key points are as follows:

- The X-machine has been shown to be Turing-complete; so can compute anything that is computable.
- It has a memory that restrains state explosion, and is the equivalent of the genetic database which informs cellular function.
- It can bi-directionally communicate with other X-machines.
- The internal functions can represent processes of any level of complexity.
- The memory contains a complete listing of the state of the X-machine at each time step, so can be used as the interface to other modeling paradigms such as a finite element solver for resolving the forces between cells.

The X-machine can be related to the cellular functions: the memory is related to gene expression; the input and output streams are signals between cell and environment (e.g. cell-cell signalling or the import/export of a growth factor molecule); the states  $Q$  could be, for instance, stages in the cell cycle; and the functions  $\phi$  define the transitions between states. These relationships are summarised in Fig. 9.3.

The classic method of communication (Kefalas et al. 2003) is to write outputs to rows of a communication matrix and read inputs from columns. This rapidly becomes impracticable in a cell model. A few  $\text{mm}^3$  of tissue contains  $O(10^6)$  cells, giving an  $O(10^{12})$  communication matrix. A simple search for neighbouring cells (as each cell communicates only with neighbours) is of  $O(n^2)$ , so equally

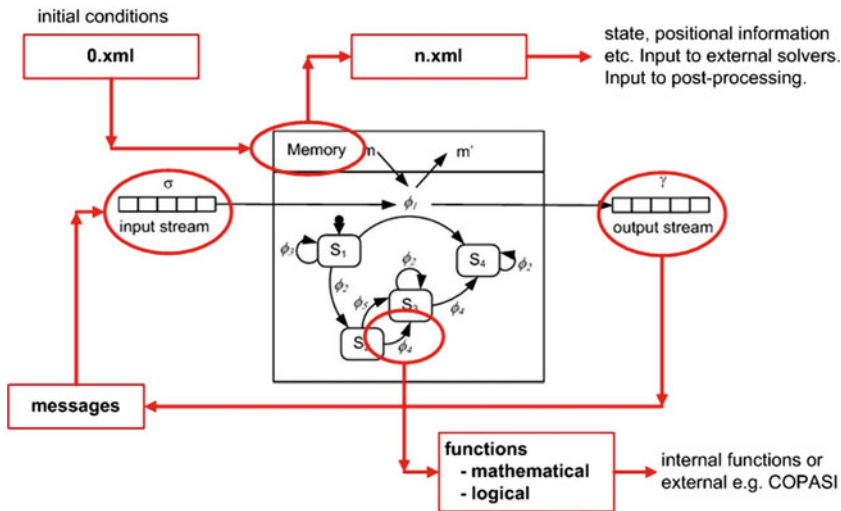


Fig. 9.3 Input and output relationships for the X-machine

intractable. There are many more efficient solutions. For spatially localised agents such as cells, the cell coordinates can be used as an index into a 3D matrix of  $O(n)$ , and the search problem is then trivial.

## 9.6 Biology, Physics, Chemistry and Computation

If the strategy for coping with cellular complexity is to start at the functional rather than the component level, a catalogue of essential and desirable biological functions is required. The over-arching aim is to develop an understanding of the development of normal structure and function in tissues from a cellular perspective, for which a representation of the cell cycle is essential (Fig. 9.4). As a minimum this has to include growth, the check points which control progression around the cell cycle, entry and exit from the  $G_0$  state (cessation and resumption of cell growth as a normal part of homeostasis, and progression to malignancy), cell division and differentiation. The behaviour of the individual cell is a function of its environment, both chemical and physical.

Externally, the physical environment involves the formation of bonds with neighbouring cells and the substrate or extracellular matrix, and responding to forces applied to the cell (mechano-transduction at focal adhesions; stretch-sensitive ion channels). The chemical environment includes receptor–ligand binding at the cell membrane (the ligands may be diffusing in the extracellular space or localised in the membrane of a neighbouring cell), and all the mechanisms for transport across the cell membrane (active and passive ion channels; importing and exporting receptors). Ion channels may communicate with neighbouring cells (gap junctions) or with the extracellular space. There are models of the internal biochemistry of



myself to more general remarks about the physical environment of the cell. My thesis is that the physical and chemical environments of the cell are of equal importance for an understanding of cellular behaviour; so any “realistic” representation of cellular behaviour has to include physics, biochemistry and the coupling between physics and biochemistry (mechano-transduction). The current situation is that physical models are considerably less well developed than biochemical models. The only area in which there has been any significant modeling of the effect of force on cellular behaviour is in mechano-electrical feedback in cardiac myocytes (Kohl and Noble 2008). The effect of individual cell behaviour (growth, division, apoptosis, motility) on tissue behaviour and the effect of tissue-level strains on individual cell behaviour have hardly been touched upon. We are working on linking individual to continuum behaviour within continuum mechanics, image analysis, signal processing and system identification (CMISS).<sup>1</sup>

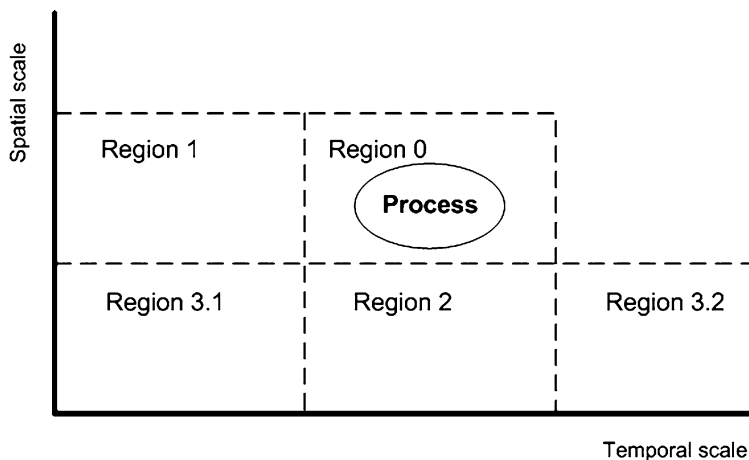
An excellent starting point for understanding the influence of subcellular components on the mechanics of a single cell is provided by Boal (2002). However, most of the cellular details have to be abstracted away to achieve a computationally tractable model containing millions of cells. The most extreme abstraction is to represent the physical cells as quasi-incompressible spheres, possibly subject to Hertzian contact mechanics, which are linked by a set of springs representing bonds and cell growth. There is a probability that cells will form bonds with each other (or the substrate) when they are sufficiently close. The simplest representation of the resulting force is a spring joining the cell centres, with a tension proportional to the separation of the cell membranes. If two cells are initially in contact, and one or both increase in size, the resulting repulsive force can again be represented by a spring joining the cell centres, with a negative tension proportional to the overlap of the membranes [the cells clearly cannot occupy the same physical space, but the separation of growth time steps (agent model) and force resolution time steps (physical model) leads to the concept of “cell overlap” being resolved by the physical model]. The detail of this simple model is provided by Adra et al. (2010) and is able to provide a plausible representation of cell movement during growth conditions which influence the formation of cell–cell bonds (Walker et al. 2004a, b, 2006a).

## 9.7 Hierarchy in Computational Models

The essence of modeling is abstraction – we would learn no more from a complete model of a cell than we would from the cell itself, and the cell computes in real time, not orders of magnitude slower. The need for abstraction is evident from a consideration of length and times scales – from  $10^{-9}$  m (molecule) to  $10^0$  m (human organism) and  $10^{-9}$  s (molecular interaction) to  $10^9$  s (human lifetime) involving  $10^5$

---

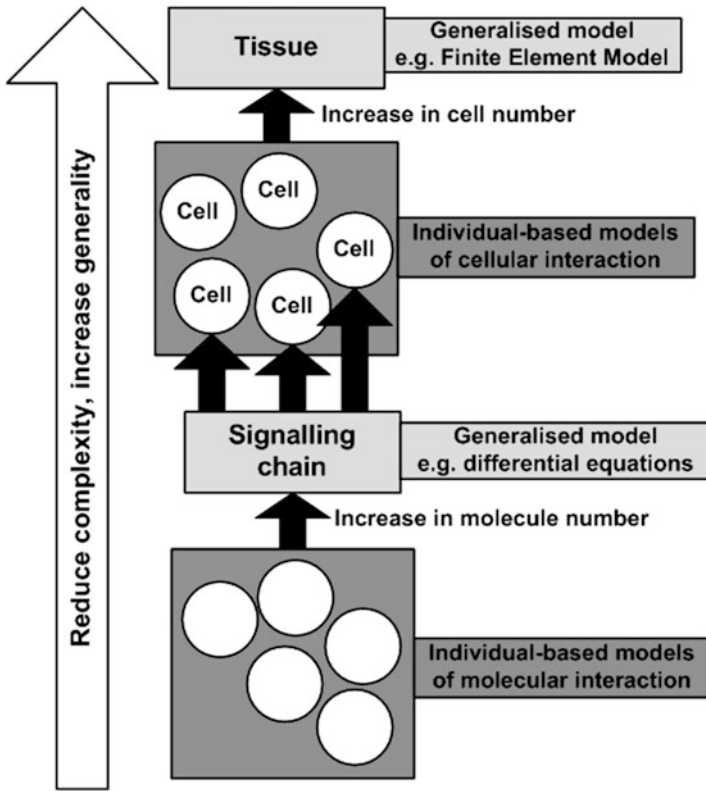
<sup>1</sup> An interactive computer program for Continuum Mechanics, Image analysis, Signal processing and System identification (CMISS), <http://www.cmiss.org>.



**Fig. 9.5** Scale separation regions

proteins. Tools are available for modeling at all of these length and time scales, but linking them to form multi-scale models is in its infancy. One promising approach is to decompose the multi-scale problem into a set of single-scale models, and then use a scale-separation map to envisage the data and control flows between the single-scale models (Hoekstra et al. 2007). For instance, the cell cycle is on a spatial scale from individual cell components ( $\sim 1 \mu\text{m}$ ) to cell size ( $\sim 10 \mu\text{m}$ ), and a temporal scale from hours to days, whereas cell signalling events are on a smaller length scale ( $< \mu\text{m}$ ) and shorter time scale (ms to hours). The relationship between scales is illustrated in Fig. 9.5.

Process A has length and time scales that place it in region 0 in the map. What is the relationship with another process B? If B also resides in region 0, there is no separation in either length or time, and a single model must be built which includes both processes. In the other cases (regions 1, 2, 3.1, 3.2) separate models are possible. In region 1, the spatial scales are the same but time scales are separated, and in region 2 the time scales are the same but the spatial scales separated. Regions 3.1 and 3.2 are separated both temporally and spatially. In 3.1, fast events on a small spatial scale are coupled to slow events on a longer scale (e.g. cell signalling coupled to the cell cycle); and in 3.2 slow events on a small spatial scale are coupled to fast events on a longer scale (e.g. cellular response to flow-induced shear stress). The scale separation map is a graph on which the vertices are the separated processes/models, and the edges are the information and control conduits between the processes. This is illustrated for the example of the vascular response to the emplacement of a coronary artery stent by Evans et al. (2008). One immediate consequence of coupling processes with differing length/time scales is that some means has to be found to reduce the amount of information that is passed from the finer to the coarser representation. One approach to abstracting information from the molecular level upwards (Fig. 9.6) is to inform generalised models from the output



**Fig. 9.6** Reducing complexity by linking individual-based and continuum models

of individual-based models – an individual-based model of molecular interaction could inform a differential equation model of cell signalling which was incorporated into individual cell models, which in turn informs a finite element model of the tissue.

A second approach, which we have adopted partly for pragmatic reasons (minimisation of effort, sharing models) and partly as this approach enables a individual cell model to be incorporated into existing physiome models such as cardiac models, is to confine the individual-based models to the cell level and link to other model paradigms (in which the required models have already been developed) at other levels. In terms of the X-machine paradigm, subcellular models (e.g. intracellular signalling or electrical excitation) are incorporated as members of the class of functions  $\Phi$ . Supracellular models (e.g. diffusion through the extracellular space or resolution of the forces acting on the cells) can be called at the end of each X-machine time step. It is usual for there to be many time steps or iterations of the linked models for each agent time step – for the EGF signalling model considered later, the diffusion time step is about four orders of magnitude shorter than the agent time step. Information exchange between agent and supracellular models is

done by reading from and writing to the X-machine memory. As far as possible, we would want to use existing models of signalling, biochemical networks and electrical activation, as building all the required models from scratch is not feasible. There are several hundred curated models available on the websites of CellML (Cell Markup Language<sup>2</sup>), SBML (Systems Biology Markup Language<sup>3</sup>) and STKE (Signal Transduction Knowledge Environment<sup>4</sup>). The key to importing these models is to incorporate an open source solver as a function call. This has been built for COPASI (Complex Pathway Simulator<sup>5</sup>), which can be used to develop ordinary differential equation models of biochemical reactions, and can import SBML and CellML models (Adra et al. 2010; Sun et al. 2010). We are currently developing a similar function call to J-Sim<sup>6</sup> which will enable the import and solution of partial differential equation models such as CellML models of electrical activation. So far, I have made the implicit assumption that the resolution of the forces between the cells is done globally, i.e. the size, position, number and type of bonds, and physical properties of each cell are passed to a global solver that resolves the forces throughout the cell mass, and returns the updated parameters to the agent model. It is also possible to resolve the forces on a local scale. Each cell only experiences the forces imposed by its nearest neighbours; so force resolution could be a function call within each cell. A small scale test of this with 2,000 cells in a monolayer (Hose, personal communication) demonstrated that two passes over the whole cell mass would resolve the forces. For the particular physical model used, the local solution was about a factor of ten slower than the global solution. However, the local solution scales linearly whereas the global solution scales as  $n^2$ ; so for a typical simulation with  $\sim 10^4$  cells the local solution would be much more efficient. An additional advantage is that, as it is embedded in the X-machine, it is inherently parallel. Nevertheless, we have chosen to not pursue this route because we see considerable advantages in being able to populate the properties of finite element representations of tissue properties in an existing package from the cellular level, and are pursuing X-machine to finite element integration in CMISS (see footnote 1), which is becoming a de facto standard for use in Physiome projects.

## 9.8 Examples at Molecular and Cell Level

As examples, I will discuss three projects. Several papers have been published on these three examples; so I will not describe the models or results in detail, but will use them to illustrate the more general discussion above. The three examples are

---

<sup>2</sup> <http://www.cellml.org>.

<sup>3</sup> <http://sbml.org>.

<sup>4</sup> <http://stke.sciencemag.org/>.

<sup>5</sup> <http://www.copasi.org>.

<sup>6</sup> <http://j-sim.org>.

NF- $\kappa$ B signalling (individual-based modeling of molecular interaction); the growth of a monolayer of epithelial cells which includes signalling, diffusion and the resolution of physical forces; and the growth of multilayered epithelial cells and the generation and healing of a wound.

### 9.8.1 NF- $\kappa$ B Signalling

This is, by a considerable margin, the simplest of the examples, as the molecules do not have any internal mechanisms and their physical interactions are ignored. Part of the signalling chain was modeled (Pogson et al. 2006). NF- $\kappa$ B (a protein) is held inactive in the cytoplasm by an inhibitor I $\kappa$ B – the two molecules are bound together to form a complex. If signalling is initiated from outside the cell (there are three mechanisms to do this), an enzyme IKK (I $\kappa$ B-kinase) starts to degrade I $\kappa$ B, releasing NF- $\kappa$ B, which can then be transported to the nucleus. Within the nucleus, NF- $\kappa$ B activates genes which control the production of I $\kappa$ B. The I $\kappa$ B is returned to the cytoplasm and inactivates NF- $\kappa$ B. This is clearly a feedback system with a time delay, and therefore has the potential to oscillate, as has been demonstrated (Nelson et al. 2004). The starting point for the individual-based model is the individual molecule (various receptors were also modeled, but the principles are the same and so they will be ignored).

If we start with a simple chemical reaction:



we can write down the differential equations and solve them. In the individual-based model, we populate the reaction vessel with a known number of molecules  $A$  and  $B$  which perform random walks (Brownian motion) throughout the vessel, interact and have a probability of forming the product  $C$ . Similarly,  $C$  has a probability of dissociating. However, if our molecules are point objects, they have an infinitesimal chance of interacting; so we assign a pseudo-volume to the molecules which defines the interaction probability (analogous to the collision cross-section in nuclear physics). The pseudo-volume is clearly related to the forward reaction constant  $k_1$ , and it is shown by Pogson et al. (2006) that the radius of the reaction volume is given by:

$$r = \sqrt[3]{\frac{3k\Delta t}{4\pi 10^3 L}}$$

where  $k$  is the reaction rate,  $\Delta t$  the time step and  $L$  the Avogadro number. The data against which the model was compared were derived from measurements on single cells in which the proteins of interest had been labelled with fluorescent tags, so that the dynamics could be quantified. The model behaviour was incorrect when

the experimental ratio of NF- $\kappa$ B to I $\kappa$ B was used. It was suggested that there was a mechanism for sequestering the inhibitor I $\kappa$ B within the cytoplasm, and that this could be binding to the cytoskeleton. A rudimentary cytoskeleton was added to the model together with a mechanism for reversibly binding I $\kappa$ B and actin (Pogson et al. 2008). The model then gave the correct dynamics with the correct total I $\kappa$ B to NF- $\kappa$ B ratio, and the proportion of bound I $\kappa$ B was subsequently confirmed by wet biology experiment.

This model begins to address the issue of locality in biological models by spatially locating all of the agents. In particular, TIR receptors and nuclear importing and exporting receptors were localised in two concentric spherical shells representing the cell membrane and the nuclear membranes, respectively. It has been demonstrated (unpublished work) that transforming the spherical cell to a more realistic shape (determined from the confocal microscopy) does not alter the dynamics. The actin cytoskeleton was modeled as a random mesh of filaments, and all of the proteins were localised in space. This only partially addresses localisation – reactions involving receptors and cytoskeleton could only take place at membrane and cytoskeleton positions, but the reactions in the cytoplasm and nucleus took place in what was effectively a well-stirred but dilute solution. The individual-based model appears to be inherently robust, in which it is impossible to achieve non-physical results such as negative concentrations, which can be achieved by an inappropriate choice of solver for differential equations (one could argue that this demonstrates a lack of competence!). However, although localisation would appear to be a compelling argument for the use of an individual-based model of cell signalling, and the biological confirmation of the sequestration of I $\kappa$ B supports this, I am not aware of any demonstration that an approach that allows localisation of reactions has yielded important new results.

### 9.8.2 *Urothelium Monolayer Growth*

The urothelium is the epithelial tissue which lines the urinary bladder. In vitro, the cells are grown in an environment which results in a contiguous monolayer of cells. There is only one type of cell. The initial model (Walker et al. 2004a) explored the growth of cells with differing concentrations of external calcium – a low calcium regime in which the binding protein, E-cadherin, is not expressed, so that the cells do not form bonds with each other; and a physiological calcium regime in which E-cadherin is expressed so that the cells do form bonds. The individual cells possessed a cell cycle, increased in size throughout the cell cycle, and a simple physical model was included. If, during the time step of one agent, the growth of a cell resulted in its overlapping with another cell, a repulsive force was generated proportional to the amount of overlap. If a cell formed a bond with another cell (the probability of bonding depended on the distance between cells, and followed a sigmoid curve), an attractive force is generated proportional to the distance apart from the cells. The

size, position and number of bonds for each cell was exported to an external solver, the forces resolved and the new size, and position of the cells was passed back to the agent model.

The growth curves generated by this model were similar to the *in vitro* growth curves, and the final cell density also varied with exogenous calcium in a similar manner to the *in vitro cell* density. The model did not correctly reproduce the differences in growth rate with different levels of exogenous calcium, and it was postulated that this was a result of the effect of epidermal growth factor (EGF), which was not included in the model. Cells which are growing both produce their own EGFR-binding ligands and respond to exogenous EGF, which diffuses through the extracellular space, generating variations in concentration which depend on local cell activity. Incorporation of EGFR signalling therefore requires the addition of two new models: a partial differential equation model of diffusion of EGF through the extracellular space; and an ordinary differential equation model of the binding of ligands to receptors at the cell surface, and the turnover of the receptors. There are thus two global models – diffusion and force – and two local models – agent and receptor binding and trafficking. The resulting model predicts the role of EGF in cell growth *in vitro* (Walker et al. 2006a).

The case for the use of individual-based models of cells is far more compelling than for molecules, and there are many examples of cellular automata and agent-based models, and the influential Cellular Potts model is also an individual-based model (see Smallwood (2009) for details). The development of structure and function in tissues is an emergent property of the behaviour of individual cells, and it is difficult to see how emergence could be captured other than by considering the cells as individuals. The division essentially is between processes or scales in which cell growth, division and death are important, and those in which they are not. Examples of the former are developmental biology, wound healing and the mechanisms which regulate cell number. An example of the latter is the coupled electrical and mechanical activity of the ventricular wall, where the effect of changes at a cellular or subcellular level (e.g. mutations in ion channels) is important, but can be considered as affecting tissue properties and not the cellular interaction *per se*.

### 9.8.3 *Epidermis Multilayer Growth*

The epidermis is the outer layer of the skin and contains three different cell types – keratinocytes, fibroblasts and melanocytes. We have developed computational models of monocultures of keratinocytes and co-cultures of keratinocytes and fibroblasts, and have compared the behaviour of the computational models with *in vitro* monocultures and co-cultures (Sun et al. 2007, 2008). A major interest in the case of skin is the response of the cells to wounding. *In vivo*, a skin wound which is more than about 2 cm diameter will not heal. *In vitro*, the model to explore this is a scratch wound – cells are grown to confluence, a pipette is drawn across the

dish to remove a strip of cells, and the resulting behaviour of the cells is monitored. Transforming growth factor (TGF)- $\beta$  is involved in the control of differentiation and proliferation in most cells, and has other functions as well, and there are contradictory results in the literature. In order to explore these effects, a TGF- $\beta$  signalling model has been included in the epidermal model using the Copasi function call in the X-machine. The resulting computational model has been used to grow the cells to produce a full-thickness, properly structured and differentiated epidermis, a section of the cells and the basement laminar have been deleted to emulate the process of producing a scratch wound, and the re-epithelialisation of the wound has been followed. A narrow wound will heal, giving a properly structured tissue, but a wide wound fails to heal (Adra et al. 2010; Sun et al. 2010).

The skin model introduces more than one family of complex agents to represent the different cell types (the agents in the signalling model are simple, as they are the lowest level in the hierarchy). The agents of course have the same structure, but the memory, functions and states may differ for different types of cells. This is an important point – the agent model of the cell is generic and could therefore in principle represent any type of cell, including totipotent or pluripotent stem cells. The challenge is to develop the agent model of the cell in such a way that it retains this versatility.

## 9.9 A Framework for Multi-Scale Modeling

A software environment, flexible large-scale agent modeling environment (FLAME<sup>7</sup>), has been developed for the building and operation of individual-based models, with the ability to link to other modeling paradigms, and is freely available to academic users from the FLAME website (see footnote 7). In addition to cell modeling, the environment has been used for social insect modeling and also for financial modeling. The models are specified in a markup language, XXML (X-machine Markup Language). The functions are written in C, and the initial conditions are also an XML file, which is the initial memory content of the X-machines. At each step, the memory is written to an XML file, and the series of XML files contain a complete history of the model run (Fig. 9.7). A parser is available to generate C code from the XXML model description, the functions file and library files. As MPI is used for the message passing between agents, the realisation is inherently parallel and can be compiled for Windows and Linux environments, to run on stand-alone, vector, parallel and grid machines. More information, documentation and references are available on the FLAME website.

---

<sup>7</sup> <http://www.flame.ac.uk>.

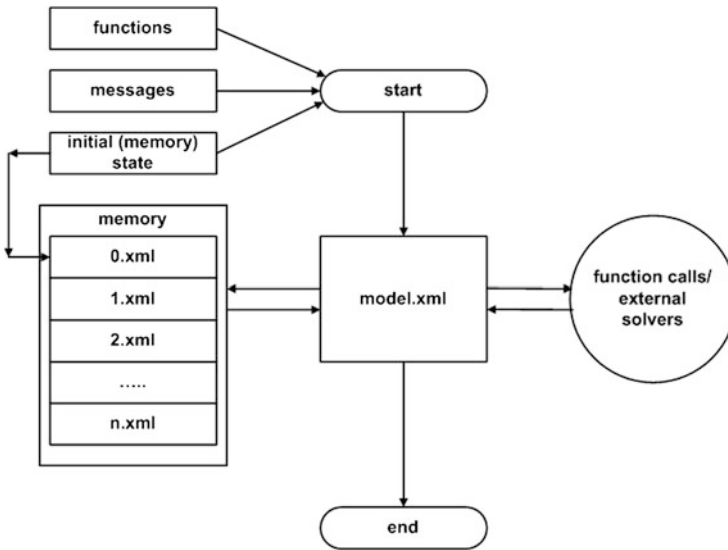


Fig. 9.7 Flexible large-scale agent modeling environment (FLAME)

## 9.10 Describing Individual-Based Models

A central feature of the scientific method is that if work cannot be repeated, it is not accepted. The implication is that the data that inform experimental work, and the tools which are used, should be freely available. This is certainly not currently the case for many of the complex computational models used in biology, and the difficulty of adequately describing complex models is a significant barrier to making information more freely available. Both CellML and SBML have published standards for model description in machine-readable formats, and maintain a large repository of curated models (i.e. models in which the equations and data have been checked, so that the model will run in a suitable software environment). There is also a web-based tool (SBML2LaTeX) available to produce human-readable text from machine-readable SBML. The situation with individual-based models is less advanced, with no widely agreed model description language and no public repository for models. Grimm (2006) have suggested a standard protocol for describing individual-based and agent-based models which they have named the Overview, Design Concepts, Details (ODD) protocol. While this is an excellent beginning, it has been formulated for ecological models, which in general appear to be very much simpler than biological models, are not multi-scale and multi-paradigm, and do not use a platform-independent description language. If we examine the model structures described above, a model description would require the following:

- A natural language description of the model [as used for the SBML repository or described by Grimm (2006)]. This should also include the data and control conduits to models which are lower and higher in the hierarchy, and between local and global models at a single level.

- A machine-readable description of the core agent model(s) – the de facto standard being a mark-up language.
- A set of initial conditions, also machine-readable.
- A set of functions (currently, for our X-machine models, these are in C, but should be in a platform-independent machine-readable format, e.g. MathML).
- A description of the interface(s) between agent model(s) and other models, including data exchange requirements and timing details.
- A machine-readable description of ordinary differential equation models (SBML or CellML) and partial differential equation models (CellML).
- A description of other model types, e.g. diffusion models and finite element models.

## 9.11 Visualisation and Graphical Output

Cell biology is a very visual discipline, with images ranging from conventional optical microscopy to electron microscopy, and a host of more recent imaging techniques such as confocal microscopy, aided by stains in the case of fixed (dead) tissue and fluorescent proteins for living tissues. Quantification includes techniques such as Western blots for dead tissue and measuring fluorescence levels for living tissue. Measurements are often made on a large number of cells to achieve the requisite sensitivity, which can mask considerable individual variation. The goal has to be to obtain time series from many individual cells (so that parameter variability can be ascertained), with multiple data points so that cellular dynamics can be compared to model dynamics. The primary target for modellers is to produce outputs that can be directly related to measures that can be made on wet biology, e.g. comparison with images or the concentrations and locations of proteins acquired from images. It is too often assumed that, because the output from the computational model looks like a biological image, the processes that produced the output must also be the same. It is a necessary, but not sufficient, condition that the outputs look the same, but it is also necessary to demonstrate that the processes are the same as well, which requires time series data.

## 9.12 Repeatability, Sensitivity Analysis and Validation

A physicist would consider most biological data to be of fairly dubious reliability – parameters have often been measured in different species and the values may vary by an order of magnitude, some parameters may be impossible to measure directly so that their magnitude has to be inferred from indirect measurements, and some may not be known at all so that an estimate (an educated guess!) has to be made. In this situation, a sensitivity analysis is essential, but for complex models the combinatorial explosion means that a complete analysis is impractical. There is

a limited literature on the validation of complex models, and the area needs more work if computational models are to be routinely used to predict the outcomes of drug trials or other interventions in disease processes. [Sornette et al. \(2007\)](#) provides a general framework for model validation with examples taken from non-biological fields, and [Marino et al. \(2008\)](#) have recently written a comprehensive account of techniques for validating computational models.

### 9.13 Lessons Learned

- Multi-scale, multi-paradigm modeling is possible.
- Logic-based (i.e. rule-based) agents are computationally lightweight and very large numbers can be modeled (equivalent to macro-scale pieces of tissue).
- Mathematical functions or more complex models will dominate the computational load.
- Encapsulation of complex models within higher level models is possible.
- Resolving forces between cells can be done, and tools to inform the constitutive equation of finite element models from individual cell models are being developed.
- There are serious validation issues which need to be addressed.
- The need for validation (in particular, for time series and robust individual cell data) is changing biological experimentation.

**Acknowledgements** The opinions expressed are my own, but have benefitted from discussions with and work by Jenny Southgate, Mike Holcombe, Sheila Mac Neil, Dawn Walker, Simon Coakley, Mark Pogson, Sun Tao, Nik Georgopoulos, Phil McMinn, Salem Adra, Des Ryan, Goodarz Kodabakshi, Rod Hose and Pat Lawford, all of whom I wish to acknowledge and thank.

### References

- S. Adra, T. Sun, S. MacNeil, M. Holcombe, and R. Smallwood. Development of a three dimensional multiscale computational model of the human epidermis. *PLoS ONE*, 5(1):e8511, 2010.
- D. Boal. *Mechanics of the Cell*. Cambridge University Press, Cambridge, 2002.
- K. Burrage, T. Tian, and P. Burrage. A multi-scaled approach for simulating chemical reaction systems. *Progress in Biophysics and Molecular Biology*, 85:217–234, 2004.
- S. Eilenberg. *Automata, Languages and Machines, Vol. A*. Academic, London, 1974.
- D. J. W. Evans, P. V. Lawford, J. Gunn, D. C. Walker, D. R. Hose, R. H. Smallwood, B. Chopard, M. Krafczyk, J. Bernsdorf, and A. Hoekstra. The application of multiscale modelling to the process of development and prevention of stenosis in a stented coronary artery. *Philosophical Transactions of the Royal Society A*, 366:3343–3360, 2008.
- E. Feytmans, D. Noble, and M. Peitsch. Genome size and numbers of biological functions. *Transactions on Computational Systems Biology*, 1:44–49, 2005.
- V. Grimm. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198:115–126, 2006.



# Chapter 10

## Interaction-Based Simulations for Integrative Spatial Systems Biology

Antoine Spicher, Olivier Michel, and Jean-Louis Giavitto

### 10.1 Introduction

It was [Fermi et al. \(1965\)](#) who proposed that computers, instead of simply performing standard calculus, could be used to study and test a physical idea. This was the introduction, in 1955, of the idea of *numerical experiments*, also called *in silico* experiments by biologists.

This epistemological and sociological change had far reaching consequences, providing to systems biology a unique tool in the investigation of biological phenomena. Computer modeling and simulation give to the biologist an access to “experimental results” that cannot be provided by direct experiments because of practical, economical or ethical reasons. However, as biologists realize the limitations of informal, intuitive analysis of complex systems ([McAdams and Shapiro 1995](#); [Von Dassow et al. 2000](#)), the computer is no longer used only to perform a computation that cannot be done analytically or by hand: its is also used to check and compare theoretical models, to systematically investigate the consequences of an hypothesis, to explore the possible range of the parameters, and to record, analyze, control and summarize some elements of the (possibly non-deterministic) behavior of a complex biological system.

Within biology, systems biology is a particularly demanding application domain since it requires to integrate several models coming from unrelated area of science like mechanics, chemistry, etc. The computer modeling and simulation of such systems require the coupling of several model fragments specifying deterministic or stochastic interactions between the system’s entities to represent continuous or discrete evolution. For instance, the modeling of the growth of the meristem at a cellular level ([Barbier de Reuille et al. 2006a](#)) requires the coupling of molecular mechanisms (e.g., chemical reaction, diffusion, active transport), mechanical stresses, developmental changes, and genetic regulation.

---

A. Spicher (✉)  
LACL – EA 4219 – Université de Paris 12, Paris EST – 61 avenue du Général de Gaulle,  
94010 Créteil Cedex, France  
e-mail: [antoine.spicher@univ-paris12.fr](mailto:antoine.spicher@univ-paris12.fr)

Computer science has developed (or appropriated) many languages and tools to help build models of real-world processes and to relate different models that operate on different levels of abstraction and various spatial and time scales. In this chapter, we advocate the use of a *rule-based framework* based on *spatial interactions* as a unifying framework for the concise and expressive simulation of a broad class of biological systems. We will address related issues such as: Can the same framework be used to model deterministic and stochastic systems? Do we need different frameworks for the expression of continuous and discrete systems? Could the same approach allow the natural and concise expression of various theoretical approaches (for the purpose of simulation)? An answer to such questions cannot be derived theoretically, but convincing elements can be provided through paradigmatic examples. This chapter is then organized as follows.

Section 10.2 discusses some of the requirements of systems biology models, the growing role of agent-based models and the current focus put on the notion of *interaction*. We emphasize also the need to handle explicit spatial relationships.

Section 10.3 presents MGS, a rule-based, spatial interaction-oriented, experimental programming language dedicated to the simulation of a broad class of biological systems.

Section 10.4 introduces the running example we use to illustrate the versatility of the rule-based approach: a synthetic multicellular bacteria or SMB. This example comes from a project presented at the International Genetically Engineered Machine Competition (iGEM)<sup>1</sup> contest in synthetic biology. SMB combines diffusion, genetic regulation, and signalling in a population.

Section 10.5 illustrates the use of the MGS rule-based approach with the development of several models of the SMB. Each model focuses on a specific time scale using a dedicated theoretical framework. We show how the MGS approach, emphasizing the notion of *spatial interactions*, is able to express concisely in the same unified and uniform simulation framework, stochastic and deterministic models, and discrete and continuous ones.

A short presentation of the perspectives and challenges opened by this work concludes this chapter.

## 10.2 Computer Modeling and Simulation in Integrative and Spatial Systems Biology

In this section, we sketch several approaches in the modeling of biological systems. We propose to base a unifying simulation framework on the spatial organization of the interaction between the entities that compose the system. An experimental programming language based on this idea is proposed in the next section and illustrated by several examples in the second part of this chapter.

---

<sup>1</sup> The SMB: Synthetic Multicellular Bacterium (iGEM'07) Paris Team Web site: <http://parts.mit.edu/igem07/index.php/Paris>.

10.2.1 Dynamical Systems in Systems Biology

Biological processes are often modeled as *dynamical systems* (Smith 1999). At any point of time, a dynamical system is characterized by a set of *state variables*. The evolution of the state over time is specified through a *transition function* which determines the next state of the system (over some time increment) as a function of its previous state and, possibly, the values of external variables (input to the system). The evolution function can be generalized to an *evolution relation* to handle non-deterministic (e.g., stochastic) evolution.

Various mathematical framework with diverse properties can be considered to formalize a dynamical system. For instance, state variables may take values from a continuous or discrete domain. Likewise, time may advance continuously or in discrete steps. Some examples of dynamical systems characterized by different combinations of these features are listed in Table 10.1. Other combinations exist and are not listed: the disintegration of a radio-active atom is a continuous-time Markov process with discrete state for instance.

These various formalisms can be applied to the same system to capture different aspects of the system’s evolution. For example, the same reaction–diffusion process (Turing 1952) in a tissue can be modeled in continuous space by partial differential equations (PDE) or in a discrete space by a system of coupled ordinary differential equations (ODE), where the state variables are the concentration of morphogens in each cell (Turing did both in his seminal paper). Reaction–diffusion processes can be also modeled by iterated mapping, sometimes called “continuous automata”, a variant of von Neumann’s *cellular automata* (CA) (Von Neumann 1966) where a cell is described by real-valued local concentrations (Turk 1991).

And totally discrete (space, time, and state) models of reaction–diffusion have also been proposed, for instance, in Greenberg and Hastings (1978).

10.2.1.1 The Need of a Unifying Simulation Language

The previous example shows that a simulation workbench for integrative biology cannot support a unique theoretical framework. In addition, even confronted to the development of one specific simulation, the programmer must cope with the wide

**Table 10.1** Formalisms used to specify dynamical systems. Some formalisms used to specify dynamical systems according to the discrete or continuous nature of time, space, and state variables. The “space” row is explained in Sect. 10.2.2.2. (C: continuous D: discrete)

| Discrete or continuous | PDE | ODE | Iterated mappings | Finite automata |
|------------------------|-----|-----|-------------------|-----------------|
| State                  | C   | C   | C                 | D               |
| Time                   | C   | C   | D                 | D               |
| Space                  | C   | D   | D                 | D               |

variety of biological entities (genes, proteins, membranes, cells, tissue, etc.). They cannot be described in a unique formalism, and yet they must be placed in a single simulation framework. This is also the case in multi-scale modeling where models of the same system at different scales can have fundamentally different characteristics (e.g., deterministic vs. stochastic).

These observations do not imply that only a general programming language can be used for the implementation of simulations in systems biology. As a matter of fact, the notion of dynamical system is a very general one, but nevertheless, it may receive some specific support that motivates the development of *domain-specific languages (DSLs)*. DSLs offer, through appropriate notations and abstractions, expressive power focused on, and usually restricted to, a particular problem domain.

We believe that it is possible to provide abstractions and notations generic enough to encompass and unify the variety of formalisms needed in systems biology. Such DSLs will support the expressive representation of various kind of states, time, and evolution functions as well as the building of coupled heterogeneous models such as discrete/continuous or stochastic/deterministic dynamical models. The programmer will be able to express the various models in a concise and expressive way, making easier to debug, tune, and evolve the simulations. Such DSL makes also possible to relate models through their implementation.

### 10.2.1.2 State and Evolution Function in Systems Biology

From the previous presentation, it is obvious that a DSL dedicated to simulation in systems biology must support in one way or the other the notions of state and evolution function. However, these two simple notions must be looked in a fresh way in the context of systems biology. Indeed, most of the systems considered in biology consist of *populations of interacting entities*. A good example is a biological cell modeled by a system of molecules that react and interact to form (other) molecules and molecular machines.

It is customary to abstract over these entities and use state variables to denote macroscopic observables or population level properties like a global concentration or a temperature.<sup>2</sup> It is assumed that as the population size increases, the behavior of the biological system is asymptotic to that of this state-variable model.

The terms *aggregate* or *mean-field* are sometimes used to qualify this approach. It allows a concise expression of the model, and despite severe limitations, mean-field approximations have been standard methodology for modeling populations of interacting entities, especially for large and homogeneous populations. One reason is that no viable alternative existed before the widespread availability of inexpensive computing power.

---

<sup>2</sup> Relying on a *mean-field* approach where the idea is to replace all interactions to any entity with an average interaction, reducing any multiple entities problem into an effective one-entity problem.

However, aggregate models rely on two assumptions that must be seriously scrutinized in systems biology:

1. The state space can be described a priori and remains fixed.
2. The global evolution function can be defined explicitly.

We examine these two assumptions in the remainder of this section.

### 10.2.1.3 Dynamical Systems with a Dynamical Structure

Very often the state space of the considered biological process cannot be described a priori. The reason is that the structure of the biological system and therefore its description (by a set of state variables) may itself vary over time, as pointed out by [Giavitto et al. \(2002a\)](#). An example is given by the development of an embryo. Initially, the state of the system is described solely by the chemical state of the egg. After several divisions, the state of the embryo is given not only by the chemical state of each cell but also by their spatial arrangement. The number of cells, their spatial organization, and their interactions evolve constantly in the course of the development and is not handled by one static structure. It means that the phase space used to characterize the structure of the state of the system at each time step must be computed jointly with the running state of the system.

The dynamicity of the structure of a biological system have been repeatedly emphasized, and several formalisms have been proposed to specify both the evolution of states and the evolution of the structure. Examples include: the concept of (*hyper*)-cycle introduced by Eigen and Schuster in the study of auto-catalytic networks ([Eigen and Schuster 1979](#)), the notion of *autopoietic systems* formulated by [Varela et al. \(1974\)](#), [Luisi \(2003\)](#), the *variable structure system* theory developed in control ([Itkis 1976](#)), or the concept of *organization* introduced by [Fontana and Buss \(1994\)](#) to formalize and study the emergence of self-maintained functional structures in a range of chemical reactions.

We call such systems *dynamical systems with a dynamical structure*<sup>3</sup> or (DS)<sup>2</sup> in short ([Giavitto and Michel 2002c, 2003](#); [Giavitto 2003](#)). Biological examples include the production of molecules and their dynamic association with multimolecular complexes ([Fontana 1992](#)) or the birth and death of cells with their mechanical constraints and signaling relations within a developing organism (morphogenesis).

### 10.2.1.4 Local Interactions

A consequence of a dynamical structure is that a global evolution function cannot be specified. As a matter of fact, if the set of variables that describe the system

---

<sup>3</sup> [Bailly and Longo \(2006\)](#) recognize the importance of this class of dynamical systems and call it “dynamicit  auto-constituante” (which could be translated to “self-producing dynamicity”), a distinctive feature of living organisms.

cannot be known in advance, it is impossible to specify a *global* evolution function (a dynamical structure is not mandatory to prevent the explicit definition of a global evolution function).

This does not mean that the (global) evolution function does not exist: it simply cannot be defined explicitly. This is the case when the individual (local) interactions between the system's entities are well characterized, but the corresponding global evolution function cannot be deduced. The macroscopic (global) evolution of the system must be computed as the "integration" of all the various local and dynamic interactions between entities.

## 10.2.2 Individual-Based Models and Their Simulations

*Individual-based models* (Lynch 2008), also called *agent-based models*, propose an alternative approach to mean-field approximation. Such models describe a system from the perspective of its constituent units and focus on the representation of the evolution of each individuals that appears in the system. As a consequence, they tackle more easily the enormous modeling difficulties raised by the dynamical structure of biological systems.

Individual-based approaches attract a renewed interest and become viable alternatives because of the increasing availability of inexpensive computing power. However, they have their own drawbacks. Their mathematical analysis appears to be at least as difficult as analysing aggregate variables models, and the simulation remains the main tool to study the system's evolution and for reaching conclusions.

Thus, beyond aggregate models, a simulation language dedicated to systems biology must be able to implement individual-based models.

### 10.2.2.1 Multi-Agent Implementation

*Multi-agent systems* (MAS) (Woolridge and Wooldridge 2001) are often advocated as the tool of choice for the implementation of individual-based models (Spicher et al. 2009). A MAS is a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions on the basis of a set of rules.

It is easy to use an agent in a MAS to represent the state of an entity part of the modeled system. The global state of the system is then the set of the state of each agent that composes the system.

However, *MAS provide no support for the notion of interaction*. Several entities are engaged simultaneously in an interaction while agents are supposed to evolve autonomously. Admittedly, in determining its evolution, an agent takes into account its neighbors. But it cannot take into account, for example, the evolution of its neighbors, which can be problematic.

A good illustration is given by a simple model of growth sometimes called the Eden model (Eden 1961). It has been used since the 1960s as a model for tumor growth. In this model, a space is partitioned in empty or occupied cells. At each step, occupied cells with an empty neighbor are selected, and the corresponding empty cell is made occupied. An exclusion principle prevents two occupied cells to invade the same empty cell.

This specification of the local evolution of the system defines the interaction between an occupied and an empty cell. It is difficult to turn this specification into a simple rule for *one* cell evolution because an empty cell can query its neighborhood to find if they are occupied cells, but they cannot know which or even if an occupied cell will invade it. Conversely, when an occupied cell decides to invade an empty one, it cannot determine whether another occupied cell makes the same decision at the same time.

### 10.2.2.2 The Spatial Structure of Interactions

Usually, only physically close entities interact because information exchange ultimately has a *local* character (e.g., transport of signaling molecules between neighboring cells). Thus the possible interactions of the entities in the system reflect the underlying physical space. The other way round, we can say that the spatial organization of the entities composing the system organizes also their interactions.

In Table 10.1, we have introduced an additional criterion, *space*, to categorize dynamical systems formalisms following the discrete or continuous setting used for the spatial organization of their entities. For instance, in a cellular automaton, entities called “cells” are organized in a regular lattice. In PDE, fields are localized in a continuous space and can extend over an entire subspace.

It is interesting to examine the case of aggregate models. For example, the aggregate model of a chemical reaction supposes that the chemical solution is well stirred and abstracts a population of molecules by a set of concentration variables. In this case, there is no need to record the position and the velocity of each molecules in the continuous underlying physical space: it is as if each molecule could interact with any other. The possibility of chemical reactions is then only constrained by the “compatibility” of the reactants and is better described by a discrete structure, the molecular interaction network. Even if this network represents functional constraints rather than constraints from the physical underlying space, it is obtained by “erasing” the localization of the molecules keeping only the possibility of interactions between different species.

Taking the previous discussions seriously pushes to make a switch from state and evolution function to individuals and interactions. In the next section, we will see how topological notions used to describe neighborhood relationships can be used to support the description of interactions in a programming language.

## 10.3 The MGS Domain-Specific Programming Language

The previous section points out several difficulties raised by the computer modeling and the simulation of biological processes in the context of integrative and spatial systems biology: the necessity to accommodate a wide range of mathematical formalisms (from continuous to discrete and from deterministic to stochastic), the dynamical structure of the system and the specification of the dynamics through local, spatially organized interactions.

To face these difficulties and to ease the building of a simulation program, we advocate a language-based approach through a *domain-specific language (DSL)* dedicated to the simulation in systems biology. DSLs are programming languages for solving problems in a particular domain. To this end, they provide abstractions and notations for the domain at hand. They are more attractive for programming in the dedicated domain than general-purpose languages because of easier programming, systematic reuse, better productivity, reliability, maintainability, and flexibility. Moreover, DSLs are usually small and more often declarative than imperative. Declarative programming focuses on *what* should be computed instead of *how* it must be done. Objects and constructions are close to the mathematical standards which enable an easier mathematical reasoning on programs. Thus, a declarative program is an executable specification not burdened by the implementation details and is close to the mathematical model.

In the rest of this section, we present such a DSL: the MGS modeling language.<sup>4</sup> MGS allows a clear and concise specification of processes through spatial interactions. In MGS, the state of a dynamical system is specified using an original and generic data structure: the *topological collection* (Giavitto and Michel 2002a). Topological collections are based on the topological relations between the interacting subparts of the system. Furthermore, the specification of the evolution law, through local interactions, is simplified by the definition of *transformations*. Transformations are functions defined by a set of rules. Topological collections and transformation are handled in a *declarative* style.

The notions of topological collection and transformation subsume several models of computation inspired by biological systems or used in their modeling and in their simulation like CA, Lindenmayer systems (used in growth plant modeling) or P systems (advocated for the modeling of compartmentalized molecular interaction networks). The MGS approach is illustrated in Sect. 10.5. The benefits of the MGS approach have been demonstrated through a number of complex applications in system biology (see Sect. 10.6).

### 10.3.1 Topological Collection

One of the key features of the MGS language is its ability to describe and manipulate a *collection of entities structured by a neighborhood relationship*. Such device,

---

<sup>4</sup> The Website of the project is <http://mgs.spatial-computing.org>.

called a *topological collection*, is used to represent the state and the organization of a biological systems: the elements of the collection are the components of the system, and the topology of the collection sets the potential interactions (i.e., two elements in the collection may interact only if they are neighbor).

Intuitively a topological collection generalizes the notion of *field* widely used in physics: each collection is build on an underlying space by associating some value with each *position* in this space. Positions can be points, but also more generally lines, surfaces, volumes, etc. The value associated with a surface may represent a flux, the value associated with a volume may represent a concentration, etc.

Topological collections can also be thought as a generalization of the notion of array where the index of an element is replaced by a position in the underlying space (Giavitto and Michel 2002a). This view subsumes a large family of important data-structures used in simulations. For instance, a labeled graph is a special case of a topological collection where the positions are the nodes of the graph and the neighborhood relationships are given by the edges of the graph.

Technically, the formalization of topological collections relies on the notion of *chain complex* (Munkres 1984) defined in algebraic topology and has been thoroughly studied in previous work of the authors (Giavitto and Michel 2002c; Giavitto and Spicher 2008b).

Several neighborhood relationships are expressible in MGS. In the rest of this chapter, we will mainly use *records*, *multiset*, *group-based fields (GBF)*, and *Delaunay* collections. These collections represent several important families of interactions and we will show that they are homogeneously handled in MGS. In addition, MGS allows heterogeneous collections (the elements of a collection can have different types) and the arbitrary nesting of collections (i.e., an element of a collection can itself be a collection). These features greatly facilitate the development of models and their simulations.

### 10.3.2 Transformation

Topological collections represent an adequate medium to specify the interactions between the elements of a biological system. In MGS, the specification of a *transformation T*:

$$\text{trans } T = \{ \quad \dots \quad \sigma \Rightarrow f(\sigma, \dots); \quad \dots \quad \}$$

corresponds to the definition of a set of rules, where the left-hand side  $\sigma$  is a pattern, matching for a subcollection, and the right-hand side  $f(\sigma, \dots)$  is an expression that evaluates a new subcollection that will be inserted in place of the matched one. The notion of subcollection depends on the neighborhood relationships of the collection: a subcollection is a connected subset of a collection and two elements are connected if they are neighbors.

Intuitively, a rule represents a possible (local) evolution of a (sub)system. The pattern in the left-hand side of the rule represents a potential configuration, and the expression in the right-hand side computes the local evolution of this configuration.

A very simple transformation is given by:

```
trans simpleT = { 0 => 1; }
```

This transformation is composed of a single rule which replaces the value 0 in the collection by the value 1. There are two important points to note.

First, this transformation may be applied to any kind of collections. Such a transformation is called *polytypic* (Jansson and Jeuring 1997). Polytypic transformations encapsulate an abstract process that can be reused in a variety of situations. For example, MGS is expressive enough to allow the definition of a generic diffusion process that can be used on any kind of collections (Giavitto and Spicher 2008b).

Second, if the transformation `simpleT` defines the replacement of 0 by 1, it does not specify which 0s must be replaced. If there are several occurrences of 0 in the collection, do we have to replace all of them, some of them, or just one of them? In the two latter cases, how are the occurrences chosen? These choices are under the control of a *rule application strategy*. The application of the transformation  $T$  on a topological collection  $e$  using a strategy  $St$  is written as:

$$T[\text{strategy} = St](e)$$

In the current implementation of MGS, all available strategies are built-in (but the functional composition of the transformations allows a certain flexibility for specific requirements). In the following, we will use two of them: the *Gillespie strategy* based on the stochastic simulation algorithm proposed by Gillespie to simulate chemical reactions (Gillespie 1977) (see Sect. 10.5.3) and the *maximal-parallel strategy* widely used in the context of L-systems (Lindenmayer 1968a) and P systems (Păun 2001). In the maximal-parallel strategy, which is the default strategy, a maximal set of non-intersecting occurrences of the pattern are simultaneously replaced by the right-hand side of the rule. When several such sets exist, one of them is non-deterministically chosen.

### 10.3.3 Two Models of Diffusion

We illustrate the notion of transformation with the simulation of a paradigmatic diffusion process. Diffusion is defined in a continuous setting in one dimension by the following equation:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2}$$

where  $D$  is the diffusion coefficient,  $u$  is the concentration of the diffusing substance, and  $x$  is the position.

Below, we describe two different approaches to simulate the diffusion of a chemical on a one-dimensional rod.

### 10.3.3.1 The Numerical Resolution of the Continuous Model

This example shows MGS' ability to handle a continuous model. By their nature, computer simulations operate in discrete time. Models initially formulated in terms of continuous time and space must therefore be discretized. Using a simple finite difference method, the previous equation is discretized as:

$$u(i, t + dt) = u(i, t) + h \sum_j (u(j, t) - u(i, t))$$

where

$u(i, t)$  represents the concentration at time  $t$  of the  $i$ th element of the discretized rod, and  $j$  ranges over the neighbor of  $i$ . Parameter  $h$  depends on the discretization and on the diffusion constant  $D$ .

This computation can be programmed in MGS by the following transformation:

```
trans diffuse[h] = {
  u => u + h * (neighborsfold(+, 0, u)
               - u * neighborsize(u))
}
```

where

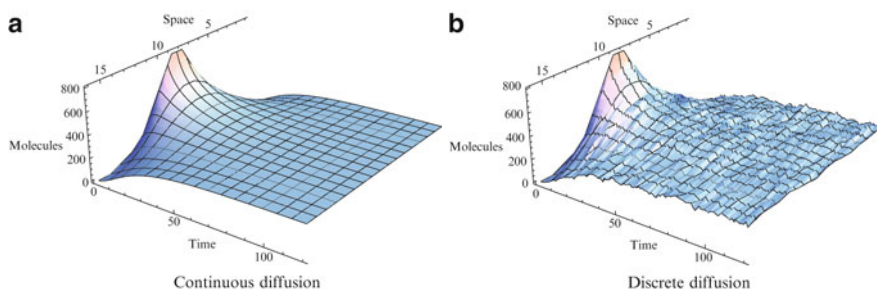
$u$  is a pattern variable that matches any element in a collection, the expression `neighborsfold( $op$ ,  $e$ ,  $u$ )` uses operation  $op$  to combine the values of the neighbors of  $u$  starting from the initial value  $e$ , and `neighborsize( $u$ )` returns the number of neighbors of  $u$ . An additional parameter is provided between brackets after the name of the transformation and corresponds to the parameter  $h$ .

It is straightforward to extend this process to a surface or a volume instead of a 1D rod. More elaborate discretization schemes are also handled similarly: for example, we give in the annex of this chapter the MGS program corresponding to the implementation of the Range–Kutta methods.

### 10.3.3.2 The Discrete Stochastic Evolution of a Diffusing Particle

Now, we want to take the same system but we focus on the level of the molecules. The rod is discretized as a sequence of small boxes, indexed by a natural integer, each containing zero or many molecules. At each time step, a molecule can choose to stay in the same box, or to jump to a neighboring box, with the same probability  $p$  (whose value depends on the time and space discretization). The state of a molecule is the index of the box where it resides. The entire state of the system is then represented as a *multiset* of indices.

A multiset is a generalization of a set (Banâtre et al. 2006): the same element can appear multiple times in a multiset. In a multiset, each element is neighbor of any others. Thus, a multiset is a good idealization of a well stirred “chemical



**Fig. 10.1** Evolution of a chemical diffusing in a 1D rod. Evolution modeled as a continuous process (a) or as a discrete stochastic one (b). Intuitively, the left figure is the limit of the right figure when the number of boxes in the rod and the number of particles grow to infinity

soup” (Giavitto et al. 2004). In our example, if there are  $m$  molecules in the box numbered  $n$ , then there is  $m$  occurrences of the integer  $n$  in the multiset.

The evolution of the system can then be specified as a transformation with three rules:

```
trans diffuse[p] = {
  q = { P = (1 - 2*p) } => q
  q = { P = p       } => q + 1
  q = { P = p       } => q - 1
}
```

The arrow construction  $=\{ \dots \} =>$  is used to specify the specific parameters of a rule. Here we give a value to the parameter  $P$  used in the probabilistic application of the rule. In this strategy, a matched pattern is replaced by the right-hand side of the rule only with a probability  $P$ . Additional rules (not shown here) are provided to deal with boundary conditions.

Figure 10.1 illustrates the iteration of the continuous and stochastic transformations. In the initial state, all particles are randomly distributed in the middle third of the rod.

## 10.4 A Synthetic Multicellular Bacterium

In the forthcoming sections, we propose to illustrate the expressiveness brought by the MGS language for the modeling, at various spatial and time scales, of the same biological process: a *synthetic multicellular bacterium* (SMB) built during the 2007 iGEM competition (footnote 1) by the French team in Paris.

We start by a short presentation of synthetic biology, the iGEM competition and then we describe the SMB project of the Paris team.

### 10.4.1 *Synthetic Biology*

Synthetic biology is an emergent field which proposes an engineering point of view on biology. It aims at building new biological systems by assembling standard low-level components called *BioBricks* (Knight 2006). These components, designed in the projects presented for the iGEM contest, are described and stored in an ontology hosted by the MIT.<sup>5</sup> They are pieces of DNA used to build biological functions (as, for example, a logical gate) and integrable within existing genomes. For example, a brick activating the production of a chemical species in the presence of a sufficient concentration of molecules of types *A* and *B* can be interpreted as a function calculating the conjunction of the chemical signals associated with the species *A* and *B*.

The basic principles of construction of the biological components, establishing the *biosynthetic* methodology, were elaborated at the MIT at the turn of the twenty-first century. They rely on classical engineering strategies: *standardization*, *decoupling*, and *abstraction* (Endy 2005). The purpose of *standardization* is twofold: to ensure compatibility between the bricks and to allow the development of generic and normalized building protocols (i.e., functioning for all bricks) economically accessible and easily implementable. *Decoupling* is a strategy that separates complicated problems into simpler ones. For instance, the separation of the various functions of a synthetic system allows the modularization of the system, the reuse of its parts, the independent evolution of each of them, etc. The separation of the phases of design and implementation reduce and eliminate the dependence between the design of a gene regulatory network and the effective building of a strand of DNA. Finally, an *abstraction hierarchy* supports the engineering of integrated genetic systems by hiding information and managing complexity through relevant levels of expression: from DNA nucleotides to parts, devices, and (complete biological) systems. Abstraction levels limit the exchange of information across levels and allow individuals to work at any level without regard for the details that define other levels.

### 10.4.2 *The International Genetically Engineered Machine Competition*

The iGEM is a competition launched by the MIT in 2003. More than 110 teams coming from all over the world participated in the 2009 issue. The competition is aimed at undergraduate students who are given the opportunity to manipulate complex molecular biology processes made simple by the synthetic biology principles. During a 3 months time period, students mentored by post-graduate students and researchers design, model, and assemble BioBricks to produce *new biological*

---

<sup>5</sup> The BioBricks are available in the *Registry of Standard Biological Parts* at the following Web site [http://partsregistry.org/Main\\_Page](http://partsregistry.org/Main_Page).

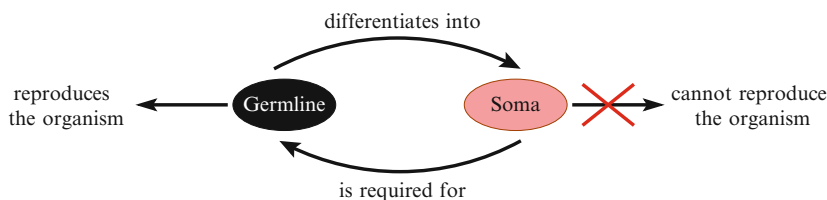
*functions* integrated into living systems. At the end of the competition, all teams gather at the MIT in the first weekend of November during the *Jamboree* where their projects are being evaluated.

In 2007, a French team supervised by A. Lindner and S. Bottani participated in the competition and was ranked first in the “foundational research” category for their *Synthetic Multicellular Bacterium* project. MGS was used to produce most of the simulations needed to validate the design (one simulation was done in MATLAB). In Sect. 10.5, we present several simulations that are inspired or extend the initial SMB simulations.

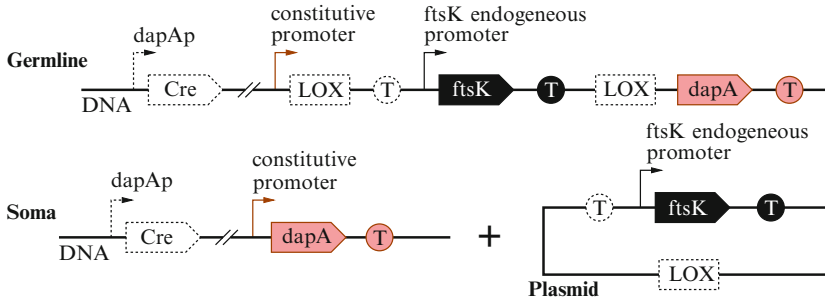
### 10.4.3 Objectives of the SMB Project

The objective of the SMB project is the design of a synthetic multicellular bacterium. This organism was thought as a tool that would allow the expression of a lethal or dangerous transgenic gene in the *Escherichia coli* bacterium without disturbing the development of its biomass. The main difficulty was to install a mechanism of irreversible bacterial differentiation which makes possible to express the transgene only in a part of the population unable to reproduce. The two lines, *germinal* (not differentiated) and *somatic* (differentiated and unable to reproduce), are interdependent and then constitute a multicellular organization (hence the name “*multicellular bacterium*”). In order to ensure that the ratio between the two populations makes it possible for the system to grow, the sterile somatic cells are designed to provide to the germinal cells a molecule essential to their reproduction: DAP (diaminopimelate). Figure 10.2 sketches the general principle of the project. Additional information is available through the iGEM Paris Team Web site (footnote 1).

The design of this organization asked for the development of two distinct biological functionalities, one for the cellular differentiation and the other for the feeding of DAP to the germinal cells. The study of this system was at the same time theoretical and practical. Although the biological implementation of the system could not be entirely carried out by lack of time, the students at iGEM Paris provided experimental evidences and theoretical proofs that the SMB organism was viable.



**Fig. 10.2** The SMB is composed of two cell types: germ cells (G) and somatic (S) cells. G cells are able to live by producing two different types of cells: G cells and S cells. S cells are derived from G cells by an irreversible differentiation step, exhibiting a new function required for the survival of the G cells. S cells cannot reproduce. This dependency between G and S cells defines the organism



**Fig. 10.3** Gene regulatory networks of the germinal and somatic cells. Gene regulatory networks describing the feeding device (light gray) and differentiation device (dotted box). Cre, dapA, and ftsK are genes, LOX is a recombination site, and T are terminators

#### 10.4.4 The Paris Team Proposal

To implement this functionality into the *E. coli* bacterium, the Paris team has proposed an original construction. The gene regulatory networks of the proposal is described in Fig. 10.3.

Two functions are described: a *feeding device* based on the production of DAP molecules (light gray) and a *differentiation device* based on a classical Cre/LOX recombination scheme (dotted box).

In the germline G, there is a natural expression of ftsK. This gene is essential for replication. The protein product of gene dapA is DAP. This protein diffuses in the environment and is rapidly degraded. However, in the germline, the dapA gene is not active since it lacks a promoter to initiate its transcription and G is auxotrophic in DAP.

The promoter dapAp is sensitive to DAP concentration. Located before the gene Cre, it allows to adjust the production of Cre to the presence of DAP in the environment. The production of Cre initiates the recombination/differentiation process.

After recombination, the genomic reassembly leads, by the excision of the parts between the two LOX recombination sites, to the cell of type S and a plasmid that is rapidly degraded. In the feeding device S, dapA is under the control of its constitutive promoter and can be expressed. The synthesized DAP diffuses in the environment allowing to reach G cells. Lacking the ftsK gene, S cells are sterile and eventually die.

### 10.5 Modeling in MGS

In this section, we illustrate the expressive power of MGS through four examples derived from the SMB. These four examples have been chosen to illustrate the MGS concepts on individual-based models as well as aggregated models, and on spatialized as well as non-spatialized models (see Table 10.2).

**Table 10.2** *Aggregated/individual-based and spatial/non-spatial models.* Aggregated models vs. individual-based models and spatialized vs. non-spatialized models in the SMB simulation examples

|                       | Aggregated                                  | Individual-based                                              |
|-----------------------|---------------------------------------------|---------------------------------------------------------------|
| Non-spatialized model | ODE<br>(Sect. 10.5.1)                       | Stochastic simulation<br><i>à la</i> Gillespie (Sect. 10.5.3) |
| Spatialized model     | Discrete diffusion of<br>DAP (Sect. 10.5.2) | Cell–cell dynamical<br>Interaction (Sect. 10.5.4)             |

### 10.5.1 Solving Differential Equations

This first modeling of SMB is a kind of proof of concept based on the study of a differential equations system. We propose here a rule-based expression of this model with two simple resolution schemes: the Euler and Runge–Kutta methods.

#### 10.5.1.1 The SMB Proof of Concept

The very design of SMB is based on the composition of a feeding device together with a differentiation device. We wonder here whether this architecture could reach homeostasis, no matter how these devices are implemented. So a minimal model is required to give such a proof of concept of the design.

To answer this fundamental question, the Paris team proposed a theoretical study of the population dynamics based on a classical differential equations model. Let  $[G]$ ,  $[S]$ , and  $[D]$  denote the concentration of germinal cells, somatic cells, and DAP molecules in a well-mixed solution. Their dynamics are captured by the three following equations:

$$\frac{d[G]}{dt} = \alpha_1 \frac{[D]^n}{[D]^n + k^n} [G] - \alpha_2 [G] - \alpha_3 [G] \quad (10.1)$$

$$\frac{d[S]}{dt} = \alpha_2 [G] - \alpha_4 [S] \quad (10.2)$$

$$\frac{d[D]}{dt} = \alpha_5 [S] - \alpha_6 [D] \quad (10.3)$$

They give the time variation of each concentration as functions of  $[G]$ ,  $[S]$ , and  $[D]$ . Parameter  $\alpha_1$  denotes the growth rate of germ cells, parameter  $\alpha_2$  denotes the differentiation rate, parameter  $\alpha_3$  denotes the death rate of germ cells, parameter  $\alpha_4$  denotes the death rate of somatic cells, parameter  $\alpha_5$  denotes the production rate of DAP by the somatic population, and parameter  $\alpha_6$  denotes the degradation rate

of DAP. In this model, the differentiation device is parameterized by  $\alpha_2$ , and the feeding device is captured by parameters  $\alpha_5$  for the DAP production and  $\alpha_1$  that is weighted by a Michaelis–Menten function representing the dependence of germinal cells growth to the DAP concentration.

### 10.5.1.2 Analysis of the ODE Model

In general, such models based on differential equations are not easily investigated. The parameters are often numerous and qualitative analyses are difficult. In our case, parameters  $\alpha_5$  and  $\alpha_6$  can be dropped assuming that the DAP concentration is stabilized (i.e., when  $[D]$  remains constant and (10.3) vanishes). This simplification of the model allows to stress out two main population behaviors. Indeed, it reveals a nontrivial fixed point  $([G]_0, [S]_0)$  that is unstable:

- For greater values of cell concentrations, an exponential growth is observed.
- For lower values of cell concentrations, both populations collapse to reach the second and trivial fixed point  $(0, 0)$ .

But is this result relevant? In other words, is the DAP stabilization assumption realistic? Should the production of DAP fluctuate, the previous sketch does not give any information on the viability of the SMB. In the following sections, we propose to focus on this question relying on different characterizations of the dynamics using numerical simulations.

### 10.5.1.3 A Numerical Solution of Differential Equations

By their nature, simulations operate in discrete time. Models initially formulated in terms of continuous time must therefore be discretized. Strategies for discretizing time in a manner leading to efficient simulations have extensively been studied. Here we use as an example a straightforward and very simple approach, the Euler method. This method particularly fits well the simulation of problems of the form:

$$\frac{d\mathbf{X}(t)}{dt} = f(\mathbf{X}(t)) \quad \mathbf{X}(0) = X_0$$

where

$\mathbf{X}(t)$  is a vector of values representing the state of the system at a given time  $t$ , and  $X_0$  is the initial state. The function  $f$  computes the variation of each coordinate of  $\mathbf{X}$  at a given time  $t$ . As far as our problem is concerned, one has the state  $\mathbf{X} = ([G], [S], [D])$  and function  $f$  corresponds to the three equations (10.1), (10.2), and (10.3).

The Euler method computes a sequence of vectors  $\mathbf{X}_n$ , where  $\mathbf{X}_0 = X_0$  at the initial time and the generic term is given by the first two terms of the Taylor expansion:

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \Delta t f(\mathbf{X}_n)$$

where

$\Delta t$  denotes the simulation time step.

We start the MGS expression of this computation by representing the state of the system in terms of topological collection. We use here a *record*. A record is one of the simplest collection consisting of two or more values so that each component (called a *field* or *member* of the record) can be accessed through a symbolic name. Each value in the record is “isolated” and has no neighbor. Hence, there is no direct interaction between the elements of a record. Elements of a record are given between braces.

The record used here has three members describing the concentrations  $[G]$ ,  $[S]$  and  $[D]$  with a value of type `float` (a real number):

```
record State = { G:float, S:float, D:float }
```

The variation of each concentration can be computed from such a state. The following function implements this procedure according to (10.1), (10.2), and (10.3):

```
fun Variation[a1,a2,a3,a4,a5,a6,k,n] (X) = {
  G = (  $\frac{X.D^n}{X.D^n + k^n}$  * a1 - a2 - a3 ) * X.G,
  S = a2 * X.G - a4 * X.S,
  D = a5 * X.S - a6 * X.D
}
```

Parameters  $a_i$ ,  $k$ , and  $n$  are given in brackets. Parameters in brackets are optional arguments. Note that the function `Variation` returns a `State`. It allows collections  $X$  and `Variation(X)` to be of the same type, and then to share the same set of positions (here fields  $G$ ,  $S$ , and  $D$ ). This property eases the computation. For example, while the concentration  $[G]$  is obtained at position  $G$  of collection  $X$  (by the expression  $X.G$ ), its variation  $\frac{d[G]}{dt}$  is at the same position  $G$  of collection `Variation(X)` (corresponding to the expression `Variation(X).G`).

Finally, one step of the Euler method can be expressed by a transformation to be applied on a collection of type `State`:

```
trans Euler[dt=Δt,f] = {
  x => let dx = f(self).(x) in x + dt * dx
}
```

In this transformation, the unique rule specifies how each element  $x$  of the collection has to be updated by computing its variation  $dx$ . This variation is taken at  $\hat{x}$  (i.e. the position of  $x$ ) of the collection and is computed by the function  $f$  (a parameter of the transformation) applied on `self`. The identifier `self` always refers to the collection which the transformation is applied to. In our example, the actual value

of `f` will be the previous function `Variation`. The whole trajectory is obtained by iterating the application of transformation `Euler` on an initial condition.

The reader is invited to note that transformation `Euler` is fully independent from the specification of `State` and `Variation`, and can be used as a generic implementation of the Euler method in many different contexts. Moreover, whereas the Euler method is sufficient for the simulations described below, we would like to underline that other integration methods can also be straightforwardly implemented in `MGS`. The implementation of the Runge–Kutta method is elaborated in the Appendix of this chapter.

10.5.1.4 Interpretation of the Simulations’ Results

Numerical approaches suffer from the lack of knowledge regarding the values of parameters. Hence, we cannot rely on any quantitative information on the system. Nevertheless, experience and classical examples give us sufficient information to determine a range of possible parameters. For the sake of the simplicity, we arbitrarily set them to the intervals given in Table 10.3.

Our objective was to observe all the possible behaviors of the system for different settings of parameters (chosen in the parameters space defined by Table 10.3) and starting from a common initial state.

The protocol of our study has consisted in running 10 000 simulations of the model. Each simulation has consisted in computing the Euler trajectory of the system over 11 000 iterations with a time step equal to 0.01 (i.e., 110 arbitrary units of simulation time) starting from an initial state where only germinal cells are present with a very high concentration of DAP. In each run, parameters were randomly chosen according to the intervals given in Table 10.3.

```
Euler[dt=0.001,f=Variation]
  ({ G = 100, S = 0, DAP = 10000 })
```

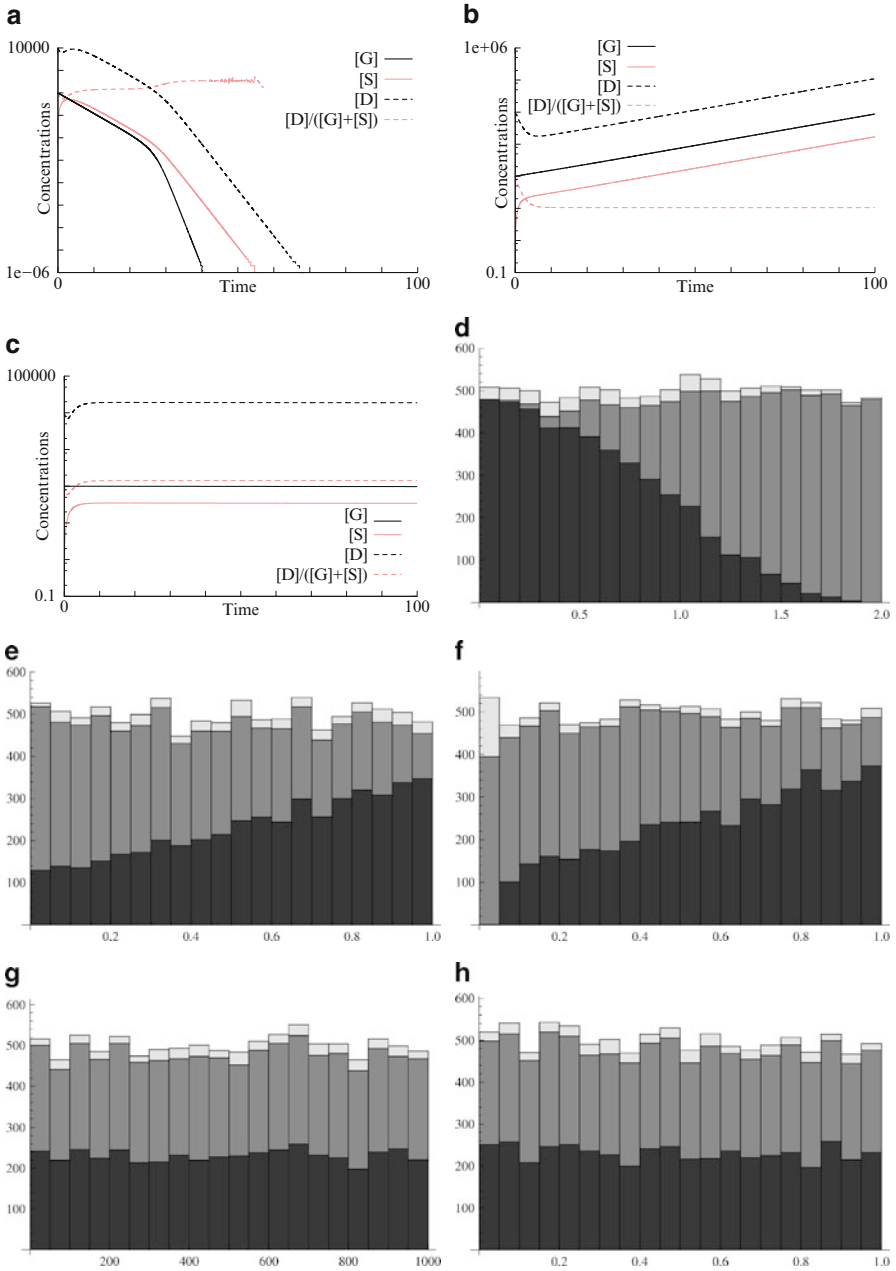
Results are given in Fig. 10.4. Only three clearly distinguished behaviors are observed and coincide with the dynamics provided by the qualitative analysis:

- 1. Population collapse (see Fig. 10.4a).
- 2. Exponential growth of the population (see Fig. 10.4b).
- 3. The unstable fixed point (see Fig. 10.4c).

In all behaviors, the system starts by consuming DAP molecules to replicate. Once DAP concentration falls below a certain level, differentiated cells start to appear and initiate the production of DAP.

Table 10.3 Intervals of the parameters for the ODE-based model

| Parameter  | Range                 | Parameter  | Range    |
|------------|-----------------------|------------|----------|
| $\alpha_1$ | [0,2]                 | $\alpha_5$ | [0,1000] |
| $\alpha_2$ | [0,1]                 | $\alpha_6$ | [0,1]    |
| $\alpha_3$ | [0,1]                 | $n$        | 2        |
| $\alpha_4$ | $\alpha_4 = \alpha_3$ | $h$        | 100      |



**Fig. 10.4** Results of simulations of the ODE-based model. Figure 10.4a–c illustrates the three observed behaviors (resp. population collapse, exponential growth, and unclassified behavior). Figure 10.4d–h gives on stacked histograms the distribution of the 10 000 simulation runs for each parameter (resp.  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ,  $\alpha_5$ ,  $\alpha_6$ ) with populations collapse in *medium gray*, exponential growth in *dark grey*, and unclassified behavior in *light gray*

In order to understand what characteristics prevent an exponential growth, the simulations have been classified according to the three behaviors, and their distributions have been analyzed (see Figs. 10.4d–h). Each histogram shows the behavior of the system following one parameter and irrespectively of the other ones. Population collapse occurs when germinal replication rate is low (either because of a small growth rate  $\alpha_1$  or because of a high death rate  $\alpha_3$ , see Figs. 10.4d,f) or when the differentiation rate is too low (see Fig. 10.4e). The last two figures (Figs. 10.4g–h) show that the system is not perturbed by the behavior of DAP production or degradation. This explains why no additional behaviors are observed compared to the qualitative analysis. In fact, the dotted lightgray curves of Figs. 10.4a–c (that correspond to the ratio  $\frac{[D]}{[G]+[S]}$ ) show that the normalized DAP concentration remains constant after a transient phase; the assumption  $\frac{d}{dt} \left( \frac{[D]}{[G]+[S]} \right) = 0$  seems appropriate for a qualitative analysis.

The conclusion of this study is twofold:

1. The choice of DAP as the main molecule to design the feeding device is good if the auxotroph germ line is robust and grows well in the presence of DAP.
2. The differentiation device is required to be efficient (a reversible differentiation should be prohibited).

We aimed here at illustrating how MGS can be used as a prototyping tool for providing quick results and orienting further investigations. This preliminary work could be improved by taking into account more realistic parameter ranges provided by the literature. The use of data analysis techniques could also give a better knowledge (e.g., parameters correlations) on the results of the simulations.

## 10.5.2 Cellular Automata

In this second modeling, we focus on the effects on the SMB due to the spatial organization of the SMB. We propose to design a cellular automaton and to implement it in MGS.

### 10.5.2.1 The Spatial Organization of the SMB

The ODE-based model proposed in Sect. 10.5.1 considers the SMB as a molecular solution of three different species uniformly distributed in space. In reality, the system consists of two populations of cells that will be organized in space. Such an organization may induce heterogeneity in the cells distribution leading to some spatial artefacts not taken into account by the ODE model (Durrett and Levin 1994). Some interesting spatial self-organizations could even be observed: for instance, one can easily imagine that the SMB collapses at some locations while it grows

exponentially at others making some patterns appear at the population scale (Shnerb et al. 2000). As a consequence, one has to investigate whether space matters or not in the SMB development.

### 10.5.2.2 A Discrete Spatial Framework

Different formalisms allow to take space into account. A first direction consists in extending the ODE of Sect. 10.5.1 by considering the spatial distribution of concentrations (i.e.,  $[G]$ ,  $[S]$ , and  $[D]$  would depend on time but also on space). This extension would introduce in the formula the use of two additional terms to deal with the spatial diffusion of concentration of bacteria and DAP molecules. These modifications make the original ODE system become a PDE system and increase the parameter space. The associated phase space becomes more difficult to study.

CA and multi-agents systems (MAS) are another class of formalisms that explicitly consider spatial organization. Both rely on an individual-based point of view. We focus here on a CA approach.

A cellular automaton is a regular lattice of places called “cells”, where each cell is characterized by a state taken from a finite set. The global evolution of the CA consists in applying synchronously, on each cell, a local evolution function that computes the new state of the cell as a function of its current state and of the states of the cells in its neighborhood.

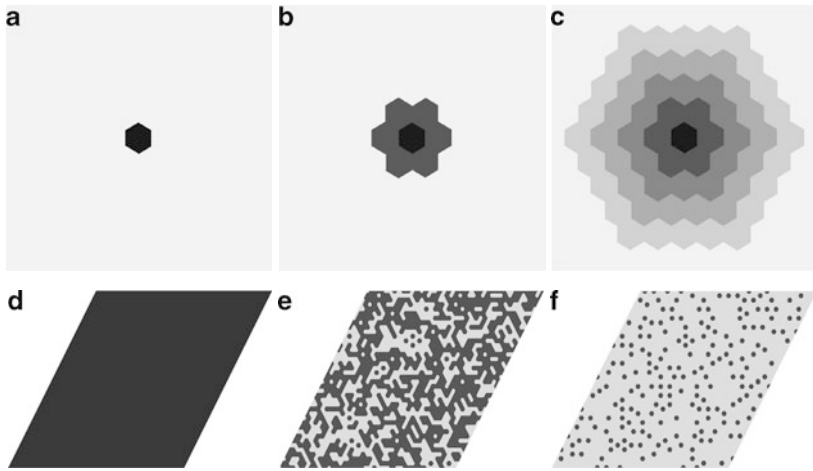
The Paris team proposed a CA to study the relation between DAP diffusion and differentiation. Their model is based on states encoded as real numbers to represent the concentration of DAP in each CA cell, and the use of non-deterministic rules using random number generators. On the contrary, we propose a totally deterministic CA with discrete states and very simple rules. More precisely, we consider a superposition of two CA: the one deals with the DAP diffusions process while the other takes into account the differentiation of the cells.

*DAP Diffusion CA.* Contrarily to the simulations of diffusion given in Sect. 10.3.3, we aim at specifying a phenomenological diffusion in a CA, that is the propagation of an information (e.g., “contains some DAP”) from a source CA cell to its neighborhood. This behavior corresponds to a classical propagation rule (Wolfram 1986) where a cell becomes activated if one of its neighbors is activated. In order to limit the radius of the propagation, the following rule may be used:

$$x_{t+1} = \max(0, y_t^1 - 1, \dots, y_t^n - 1) \quad (10.4)$$

where

$x$  denotes a cell of the CA,  $x_t$  its state at time  $t$ , and  $y^1, \dots, y^n$  the neighbors of  $x$ . Here states are encoded by integers that are gradually decremented from the source: 0 means no activation and  $n > 0$  means that the activation propagates with a radius  $n$  around the cell. Some evaporation may be introduced to deal with the removal of



**Fig. 10.5** Results of simulations of the CA model. Top line shows the propagation of DAP around an isolated source with radius 5: from left to right, initial state, state after 1 iteration, steady state. A light gray cell means no DAP, the gray scale represents the concentration of DAP and the black cell is the source. Bottom line shows the evolution of the CA defined by (10.6), on a  $40 \times 40$  hexagonal grid only filled by germinal cells with a randomly chosen ratio: from left to right, initial state, state after 13 iterations, state at fixed point. Germinal cells figure in dark gray and somatic cells in light gray

the source. A source maintains its state to a specific integer value denoted by  $N_{\mathcal{R}}$ . Figure 10.5a–c shows the discrete diffusion around an isolated source for  $N_{\mathcal{R}} = 5$  on an hexagonal grid.

**Differentiation CA.** This CA focuses on the bacterial layer. Each cell of the CA represents a part of the whole population. Under some conditions on the DAP level, a CA cell progressively goes from a majority of germinal bacteria to a majority of somatic bacteria. We abstract the ratio between the two populations in a CA cell by an integer between 0 and  $N_{\mathcal{P}}$ : 0 means that only somatic bacteria are present and  $N_{\mathcal{P}}$  only germinal bacteria are present.

The dynamics of a CA cell is as follows: if there is enough DAP, the number of germinal bacteria becomes maximal. Otherwise, this number decreases at each time step. Finally, if no germinal bacterium remains, the ratio is locked to 0. Equation (10.5) summarizes this behavior:

$$u_{t+1} = \begin{cases} N_{\mathcal{P}} & \text{if there is enough DAP} \\ 0 & \text{if } u_t = 0 \\ u_t - 1 & \text{otherwise} \end{cases} \quad (10.5)$$

where

$u$  denotes a CA cell and  $u_t$  its state at time  $t$ .

*Coupling both CAs.* Equations (10.4) and (10.5) are combined to define the final CA. Let  $c_t = (u_t, x_t)$  denotes the state of a CA cell  $c$  a time  $t$ . The local evolution function is given by:

$$c_{t+1} = \begin{cases} (0, N_{\mathcal{R}}) & \text{if } x_t = 0 \\ (u_{t+1}, \max(0, x_{t+1} - N_{\mathcal{C}})) & \text{otherwise} \end{cases} \quad (10.6)$$

where

$x_{t+1}$  and  $u_{t+1}$  are given by (10.4) and (10.5), and  $N_{\mathcal{C}}$  represents the DAP consumption of germinal bacteria. Finally, we consider that there is not *enough* DAP when  $x_{t+1} - N_{\mathcal{C}}$  is negative and no DAP source is available in its neighborhood.

### 10.5.2.3 MGS Expression of a Cellular Automaton

MGS allows an easy specification of CA. We use the *group-based field* topological collections (GBF) to represent regular and uniform lattice, as used in CA or for the numerical solutions of PDEs. The neighborhood relationships of a GBF are described in terms of a mathematical group, the group of elementary displacements in the lattice (Giavitto and Michel 2001; Giavitto et al. 2002b). The corresponding space can be visualized as a graph, the Cayley graphs of the presentation of the group (Giavitto and Michel 2002b). This abstract approach enables the easy specification and a uniform handling of a large family of circular and screwed, bounded or unbounded grids in any dimension.

For the purpose of this example, each GBF position is labelled by an MGS record of type  $\{x:\text{int}, u:\text{int}\}$  representing the state  $c^t$ . Each GBF position has six neighbors achieving an hexagonal grid.

The dynamics of (10.6) are implemented as follows:

```
trans SMB_CA = {
  c / c.u == 0 => { x = NR,          u = 0          }
  c / NoPrs(c) => { x = Diff(c) - NC, u = ND        }
  c / c.u == 1 => { x = NR,          u = 0          }
  c              => { x = 0,          u = c.u-1      }
}
```

where

`NoPrs(c)` computes whether there is not enough DAP on cell  $c$  and `Diff(c)` computes the diffusion on cell  $c$  with respect to (10.4). Note that the order of the rules matters: for instance, the matching of a cell by the third rule implies that it cannot be matched by the first two ones. This transformation is applied using the standard maximal parallel strategy of MGS.

### 10.5.2.4 Interpretation of the Simulations' Results

Figure 10.5d–f shows how differentiation appears in a population of germinal cells. As the CA transition function is deterministic, symmetry is broken in the initial state (otherwise all cells would exhibit the same behavior): we have chosen to start the simulation with cells of the form  $\{ \mathbf{x} = 0, \mathbf{u} = 1 + \text{random}(N_{\mathcal{R}}) \}$ , that is to say germinal cells with a ratio uniformly chosen in  $[1, N_{\mathcal{R}}]$ , and no DAP.

No symmetrical pattern appears during the simulation, whatever the parameters  $N_{\mathcal{R}}$ ,  $N_{\mathcal{P}}$ , and  $N_{\mathcal{C}}$  are. The distribution of the differentiation cells follows the ratio distribution chosen at the initial state. An equivalent behavior is observed when the symmetry is broken by randomly initializing the field  $\mathbf{x}$ . The uniformity of the dynamics supports the assumption of a well-mixed solution used in Sect. 10.5.1 and confirms the previous result.

We have shown with this model how a rule-based programming style fits well the specification of CA. Not more than ten lines are required to describe it in MGS. Moreover, thanks to the polytypic feature of MGS, the specification of the topology is totally decoupled from the definition of the dynamics; transformation `SMB_AC` could be applied on any kind of topological collection, and more especially on any kind of grids and neighborhoods (like square grids with von Neuman or Moore neighborhoods, toric grids, etc.). More specialized CA tools are often ad hoc and do not exhibit so much flexibility and genericity in the expression of models.

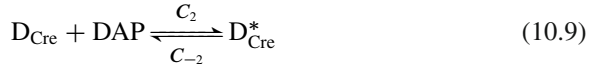
## 10.5.3 Stochastic Simulations

The two previous approaches provide results at the level of the synthetic device. In this section, we study the construction of these devices in terms of biological parts and synthetic construction as described in Sect. 10.4.1. More specifically, we propose a stochastic model of the SMB at the level of one bacterium together with its simulations using the exact *stochastic simulation algorithm* defined by Gillespie (1977).

### 10.5.3.1 Robustness Analysis of the SMB Design

The characterization of a synthetic device depends on its implementation. We aim at checking if the behavior of the Paris team construction respects the main objective of the SMB. More precisely, we focus on the noise sensitivity and the relation between parts parameters (like the rate of DNA Cre/LOX recombining) and the devices parameters (such as the differentiation rate).

A common way of modeling gene regulation is to consider the regulatory network as a set of biochemical reactions. The set of chemical interactions induced by the Paris team construction (see Fig. 10.3) are abstracted by the following reactions:



These reactions involve two kinds of chemical species: the DAP and Cre molecules, and the DNA constructions of Fig. 10.3 abstracted by:

- $\text{D}_{\text{Cre}}, \text{D}_{\text{Cre}}^*$ : Differentiation-free part of the construction composed of promoter  $\text{dapAp}$  and the coding region for Cre. The two symbols represent, respectively, the activated (no DAP repression on  $\text{dapAp}$ ) and inhibited (DAP binds  $\text{dapAp}$ ) state of the promoter.
- $\text{D}_{\text{G}}, \text{D}_{\text{G}}^*, \text{D}_{\text{S}}$ : Part of the DNA modified by the Cre/LOX recombination mechanism,  $\text{D}_{\text{G}}$  before and  $\text{D}_{\text{S}}$  after recombination.  $\text{D}_{\text{G}}^*$  corresponds to an intermediate state where only one LOX site is bound by Cre.

Reactions (10.7) and (10.8) describe the natural degradation of molecules Cre and DAP. Reactions (10.9)–(10.11) express the behavior of the promoter  $\text{dapAp}$ : inhibition/activation by DAP and production of Cre (Reactions (10.10) and (10.11) differ in their reaction constants:  $C_4 \ll C_3$ ). Reactions (10.12) and (10.13) specify the two steps of a Cre/LOX recombination:  $\text{D}_{\text{G}} \rightarrow \text{D}_{\text{G}}^* \rightarrow \text{D}_{\text{S}}$ . The regulation induced by  $\text{D}_{\text{G}}$  (expression of the gene  $\text{ftsK}$ ) is not considered in this model. On the contrary, the behavior of  $\text{D}_{\text{S}}$  is specified by (10.14), corresponding to a constitutive production of DAP. The last reaction (10.15) expresses importation and exportation of DAP from the extracellular environment, where  $\text{DAP}_{\text{ex}}$  denotes the external occurrences of DAP molecules.

### 10.5.3.2 Stochastic Modeling for Sensitivity to Noise Analysis

The Paris iGEM team has chosen to investigate this kind of molecular model using a differential equation approach based on the mass action law. Thanks to this study, they provided a set of optimized parameters for an exponential growth of the SMB. Nevertheless, such results may be biased since the differential approach (1) relies on a global homogeneous assumption and (2) does not take noise into account. Since the number of molecules involved in gene regulation is in general very low, a stochastic approach may provide complementary result on noise sensitivity.

A usual abstraction in the simulation of biochemical systems consists in considering the system (here the bacterium) as an homogeneous chemical solution where the reactions of the model are taking place. Gillespie (1977) has proposed an algorithm for producing the trajectories of such a chemical system by computing the *next reaction* and the *elapsed time* since last reaction occurred. Let  $\mu$  be a chemical reaction, the probability that  $\mu$  takes place during an infinitesimal time step is proportional to:

- $c_\mu$ , the *stochastic reaction constant*<sup>6</sup> of reaction  $\mu$ .
- $h_\mu$ , the number of distinct molecular combinations that can activate reaction  $\mu$ .
- $d\tau$ , the length of the time interval.

Gillespie proved that the probability  $P(\tau, \mu)d\tau$  that the next reaction will be of type  $\mu$  and will occur in the time interval  $(t + \tau, t + \tau + d\tau)$  is:

$$P(\tau, \mu)d\tau = a_\mu e^{-a_0\tau}d\tau$$

where

$a_\mu = h_\mu c_\mu$  is called the *propensity* of reaction  $\mu$ , and  $a_0 = \sum_v a_v$  is the combined propensity of all reactions.

This probability leads to the first straightforward Gillespie's algorithm called the *first reaction method*. It consists in choosing an elapsed time  $\tau$  for each reaction  $\mu$  according to the probability  $P(\tau, \mu)$ . The reaction with the lowest elapsed time is selected and applied on the system, making its state evolve. A new probability distribution is then computed for this new state, and the process is iterated.

### 10.5.3.3 Gillespie-Based Simulations in MGS

Here, we consider the bacterium as a well-mixed chemical solution. It can be represented by a multiset, that is a topological collection, where any element may interact with all the others (Fisher et al. 2000). The simulation of "real" chemical reactions

---

<sup>6</sup> Evaluating the stochastic constants is one of the key issues in stochastic simulations of biochemical reactions. The interested reader should refer to De Cock et al. (2003) and Zhang et al. (2003) for the description of two experiences in that field.

requires a strategy for multiset rules application in accordance with the reactions kinetics. The MGS language provides such a strategy based on Gillespie's algorithm. We propose to use this strategy for simulating the previous set of chemical reactions.

As said above, the state of the bacterium is represented by a multiset of values. The considered molecules are abstractly represented using MGS *symbols* denoted by back-quoted identifiers. For example, the MGS symbol ``Cre` corresponds to one molecule of Cre. Thus, each chemical reaction is translated into a transformation rule (or two if the reaction is reversible) characterized by an arrow parameter `C` representing the stochastic constant of the reaction. For example, the reversible reaction (10.9) can be straightforwardly translated to the two following MGS rules:

$$\begin{array}{ll} \text{'dCrA, 'DAP} = \{ C = C_2 \} \Rightarrow \text{'dCrI} & \text{and} \\ \text{'dCrI} = \{ C = C_{-2} \} \Rightarrow \text{'dCrA, 'DAP} \end{array}$$

Consequently, the whole dynamics is captured by the following set of rules in transformation SMB\_STO:

```
trans SMB_STO[DAPEX] = {
  `Cre      = { C = C0      } => .
  `DAP      = { C = C1      } => .
  `dCrA, `DAP = { C = C2      } => `dCrI
  `dCrI      = { C = C-2     } => `dCrA, `DAP
  `dCrA      = { C = C3      } => `dCrA, `Cre
  `dCrI      = { C = C4      } => `dCrI, `Cre
  `dG1, `Cre  = { C = 2 * C5   } => `dG2
  `dG2, `Cre  = { C = C5      } => `dS
  `dS        = { C = C6      } => `dS, `DAP
  `DAP       = { C = Cex      } => (DAPEX++; .)
  .         = { A = DAPEX * Cim } => (DAPEX--; `DAP)
}
```

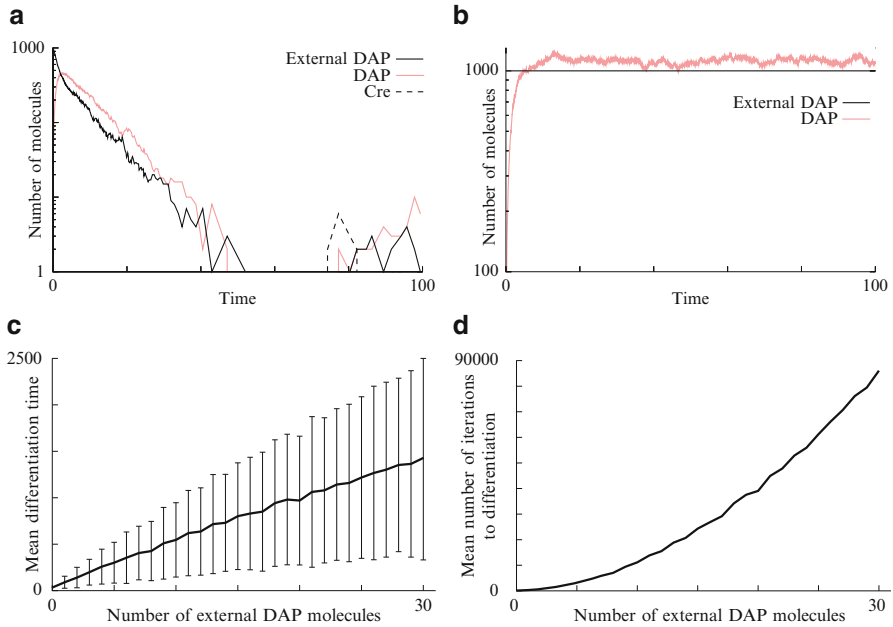
In the last two rules, the external  $\text{DAP}_{\text{ex}}$  molecules are specified by the counter `DAPEX` given as an optional parameter. This counter is incremented (resp. decremented) when a DAP molecule is imported (resp. exported).

The last rule of transformation SMB\_STO explicitly computes the propensity `A` instead of using the usual parameter `C`. This feature allows a fine control of the Gillespie application strategy when required.

### 10.5.3.4 Interpretation of the Simulations' Results

A simulation is run by calling the transformation SMB\_STO using Gillespie's strategy:

```
SMB_STO[strategy=`Gillespie, DAPEX=1000]
  (`dCrA::`dG1::():bag)
```



**Fig. 10.6** Results of simulations of the stochastic-based model. On top, two examples of stochastic simulations when external DAP (in solid black) remains constant (Fig. 10.6b) or not (Fig. 10.6a). Internal DAP is drawn in light gray and Cre in dotted line. Figure 10.6c shows the mean simulation time to differentiation with a constant external DAP concentration for different values of that concentration. Vertical bars correspond to the standard deviation. Figure 10.6d shows the mean computation time to differentiation

The initial state is specified by two molecules, namely  $D_{Cre}$  and  $D_G$ , added (with the insertion operator  $:$ ) to an empty multiset (denoted by  $()$ :`bag` in the MGS syntax). This state represents a germinal cell. External DAP is specified by initializing the counter  $DAP_{ex}$  to 1000 molecules.

Top line of Fig. 10.6 gives two different runs of the simulation corresponding to the evolution of Cre, DAP, and  $DAP_{ex}$  populations over 100 arbitrary units of time (AUT).

Figure 10.6a shows the classical behavior of a germinal cell: the  $DAP_{ex}$  is imported from the outside until no DAP remains in the system (this process takes 50 arbitrary units of time). During this first part of the simulation, the expression of Cre is repressed by the over representation of DAP. After that, the repression decreases during 20 AUT until some Cre molecules are produced. At time 80 AUT, DAP molecules appear which means that the differentiation occurred. On the contrary, Fig. 10.6b shows the evolution of a germinal cell when  $DAP_{ex}$  remains constant (i.e., expressions  $DAP_{ex}++$  and  $DAP_{ex}--$  are removed from the specification of `SMB_STO`). DAP molecules are exchanged between the cell and its environment until an equilibrium is reached. While any differentiation occurs during this simulation, the germinal cell will differentiate since parameter  $C_4$  is not null.

We propose to use the stochastic model to evaluate the differentiation rate of the SMB. More specifically, we focus on the mean simulation time required for a germinal cell to differentiate while  $\text{DAP}_{\text{ex}}$  is constant. Results are given on the bottom line of Fig. 10.6. The protocol of this experiment consists in running 1000 simulations for each value of  $\text{DAP}_{\text{ex}} \in [0, 30]$  and starting with the same initial state. Each simulation stops when the differentiation occurs (i.e., when  $\text{dS}$  appears in the collection). Mean times and associated standard deviation are plotted in Fig. 10.6c. Surprisingly both quantities seem to behave *linearly* with  $\text{DAP}_{\text{ex}}$ .

One has to pay attention to the fact that such simulations are costly in computing time. Figure 10.6d gives the mean computation time of the simulation, showing that it increases more than linearly with the value of  $\text{DAP}_{\text{ex}}$ . Actually, Gillespie's algorithm, in its original definition, only allows a small number of molecules.

As a conclusion, one can establish that the differentiation rate is easily related to the quantity of DAP released in the environment by somatic cells. Such a result is meaningful as it relates quantities of different scales: the population and cellular scale of the differentiation and the genetic and molecular scale of DAP concentration.

Our stochastic simulations exhibit low concentrations of chemicals within a cell (e.g., there are less than ten molecules of Cre after the differentiation occurs). This result questions the pertinence of ODE models such as the one proposed by the Paris team (footnote 1).

### 10.5.4 Integrative Modeling

So far, we have considered classical ways of modeling and simulating a biological process at a given level of description. In this last model, we aim at simulating the entire population of bacteria with an explicit representation of cells in a 2D space.<sup>7</sup> In addition, we want to integrate in the model physical and biological behaviors. Our proposition is based on the specification of cell–cell dynamical interactions and the computation of the neighborhood of the cells using an implicit Delaunay triangulation.

#### 10.5.4.1 Description of the Model

In our proposition, bacteria are represented by circles localized in the 2D Euclidean space, with a radius depending on their size. Bacteria push away each other and consequently change their position in space and their immediate neighborhood. Thus, this neighborhood is required to be dynamically computed according to the spatial

---

<sup>7</sup> The third dimension is not considered as the SMB is supposed to grow in the plane of a Petri dish, for example.

coordinates of their associated circles. This approach has already been successfully used in systems biology for the modeling of cell population (Gibson et al. 2006; Barbier de Reuille et al. 2006a).

The modeling of SMB is organized into two coupled models: a mechanical model and a biological model. A cell is encoded by a record which includes the position, the radius, the local DAP concentration, the differentiation state (germinal or somatic), etc. Cells are elements of a Delaunay topological collection. This kind of collection computes implicitly and transparently the Delaunay triangulation of a set of entities embedded in  $\mathbb{R}^n$  (Aurenhammer 1991).

*Mechanical model.* The mechanical model consists of a mass/spring system. Bacteria are considered as punctual masses localized at the center of their associated circle; the presence of a spring between two masses depends on the neighborhood computed by the Delaunay triangulation. The mechanical effects of the growth of the bacteria is captured by the elongation of the springs rest lengths. Thus, each cell computes its acceleration by summing all mechanical forces induced by its incident springs, and consequently moves in space. This is done by the transformation *Meca*.<sup>8</sup> *Meca* sums the forces applied on each cell using a *neighborsfold* expression. Then, the Euler transformation (see Sect. 10.5.1) is used twice to integrate during a time step  $\Delta_1 t$  acceleration into velocity and velocity into new positions.

*Biological model.* The cellular model is an extension of the CA given in Sect. 10.5.2. The discrete DAP diffusion is replaced by the classical continuous model given in Sect. 10.3.3, and a stochastic differentiation is used instead of the notion of populations ratio. New rules are added to deal with cellular growth, division, and death: in the presence of DAP, G cells grow by increasing their radius. When the G cell radius reaches a threshold, the cell divides. S cells keep on growing and then die when another threshold is reached. The corresponding transformation is called *Cell18* and computes the cellular evolution during a period  $\Delta_2 t$ .

#### 10.5.4.2 Integration of the Two Models

As classical functions, transformations can be arbitrarily composed. This is the key to the coupling of the two models. The iteration of a function can be specified by the MGS option *iter*. It allows to deal with different time scales: assuming that the mechanical process is faster than the cellular process (i.e.,  $\Delta_2 t > \Delta_1 t$ ), the whole model is captured by the following evolution function:

```
fun SMB_DEL(state) =
  Cell [dt= $\Delta_2 t$ ] (Meca [dt= $\Delta_1 t$ , iter= $\frac{\Delta_2 t}{\Delta_1 t}$ ] (state))
```

<sup>8</sup> The whole MGS program of the simulation is available at [http://mgs.spatial-computing.org/integrative\\_biology.tgz](http://mgs.spatial-computing.org/integrative_biology.tgz).

where

option `dt` corresponds to the time step used in transformations `Meca` and `Cell`.

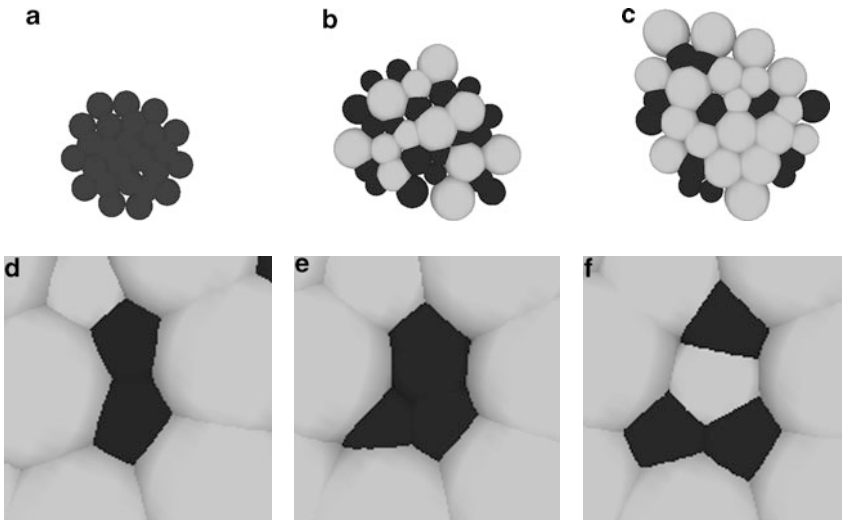
Here transformation `Meca` is applied  $\frac{\Delta_2 t}{\Delta_1 t}$  times for only one application of `Cell`.

### 10.5.4.3 Interpretation of the Simulations' Results

The protocol of the proposed simulation consists in iterating 10 000 times the function `SMB_DEL` starting from a small initial population of 25 germinal cells with a deficit in DAP. Screenshots of the simulation are shown in Fig. 10.7. The visualization of the evolution exhibits two interesting phenomena.

On small population size (i.e., at the beginning of the simulation), the ratio of the two cell lines fluctuates. Figure 10.7b presents a state of the population where most of the cells are germinal. On Fig. 10.7c, S cells predominate. The oscillations are due to delays between configurations where G cells are well fed (many S cells are present) and configurations of DAP starvation (not enough S cells are present). The fluctuations decrease as the population size increases, and the ratio globally stabilizes as predicted by the ODE-based model (see Fig. 10.4). Indeed, the spatial distribution blurs the synchronization between cells.

The population tends to spatially organize in the following way: uniformly distributed small clusters of G cells surrounded by somatic ones. Clusters remain small-sized because when their size increase (by G cell divisions), the interior cells lack DAP and differentiate. This dynamics is reminiscent of the behavior exhibited by the CA model of Sect. 10.5.2.



**Fig. 10.7** Results of the integrative model. Germinal cells are in dark gray and somatic cells in light gray. Figure 10.7a–c correspond to an initial population and its evolutions at time 43AUT and 62AUT. Figure 10.7d–f focus on a cluster of germinal cells surrounded by somatic cells. See text for explanation

## 10.6 Related Work, Conclusions, and Perspectives

In this short conclusion section, we present related work, close to our approach, used for the modeling and simulation of biological systems. Then, we conclude by assessing the importance of using a single and coherent domain-specific language for the modeling, at various spatial scales, of biological systems.

### 10.6.1 Related Work

In this chapter, we have illustrated the use of MGS, an experimental programming language that investigates the use of topological collections and transformations in the simulation of complex biological systems. Based on the notion of spatial interaction, MGS provides a unified simulation framework encompassing discrete/continuous and deterministic/stochastic formalisms. Even though MGS is a prototype in constant evolution, MGS' concepts have been validated on numerous applications: the crawling of the sperm cell of *Ascaris suum* (Spicher and Michel 2005), a simplified version of neurulation (Spicher and Michel 2007), the growth of the meristem at a cellular level (Barbier de Reuille et al. 2006a), molecular self-assembly (Giavitto and Spicher 2008a), the modeling of the paradigmatic phage lambda genetic switch (Michel et al. 2009), etc.

As a side-effect, changing the topology of a collection makes it possible for MGS to emulate some well-known computational models. Transformation on multiset is reminiscent of multiset-rewriting (Banâtre and LeMetayer 1993). Nesting multiset lead to P systems (Păun 2001), a new distributed parallel computing model based on the notion of a membrane structure. P systems are advocated for the modeling of compartmentalized molecular interaction networks. Lindenmayer systems (Lindenmayer 1968a), which loosely correspond to transformations on sequences or string rewriting, have long and successfully been used in the modeling of (DS)<sup>2</sup> in the domain of plant growth.

And, as shown in Sect. 10.5.2, it is straightforward to express CA in MGS.

### 10.6.2 Conclusions and Perspectives

Work in systems biology generally puts a considerable emphasis on the end result, the model of a complex biological system, and neglects so far the problem of *building* this model. The construction of a model and its use, e.g., for simulation, is a difficult task, and it often requires the combination of many formalisms and complementary approaches. We also want to stress the importance of dynamical structures in biological systems. This kind of dynamical systems is very challenging to model and simulate. New programming concepts must be developed to ease their modeling and simulation.

Using **MGS**’ topological collections and transformations allowed us to model our problem in the same formal framework: a mean-field approach using ODE for a quick proof of concept of the synthetic construction (Sect. 10.5.1), a discrete spatial model using CAs on various topologies allowing a finer analysis of the relations between diffusion and differentiation (Sect. 10.5.2), a robustness analysis on noise sensitivity on the **SMB** parts (Cre/LOX recombination), and devices (recombination rate) parameters using a stochastic modeling (Sect. 10.5.3). The last model integrates physical and biological constraints in a 2D model, making it possible to analyze the effects of the spatial distribution on the various possible configurations (Sect. 10.5.4).

The perspectives opened by this work are numerous. At the language level, the study of the topological collections concepts must continue with a finer study of transformation kinds. A lesson learned from the use of **MGS** by biologists is the needs of user-friendly interfaces and of graphical tools for the analysis of the simulation’s results. Another direction of developments is the coupling of **MGS** with various databases and repositories to retrieve parameters or to store and reuse fragments of models.

The development of **MGS** is part of a “grand challenge” aimed at the development of a methodological and technological framework that, once established, will enable the sharing of models, the analysis, and the derivation of predictive hypothesis from them.

**Acknowledgements** The authors would like to thank the reviewers for their valuable comments on a first version of this chapter.

We gratefully acknowledge all the people who contributed to make the first French participation in iGEM in 2007 a success: the students, D. Bikard, F. Caffin, N. Chiaruttini, T. Clozel, D. Guegan, T. Landrain, D. Puyraimond, A. Rizk, E. Shotar, G. Vieira, the instructors, F. Delaplace, S. Bottani, A. Jaramillo, A. Lindner, V. Schächter; the advisors, F. Le Fevre, M. Suarez, S. Smidtas, A. Spicher, and P. Tortosa.

Further acknowledgments are also due to J. Cohen, B. Calvez, F. Thonnerieux, C. Kodrnja, and F. Letierce who have contributed in various ways to the **MGS** project.

This research is supported in part by the the University of Évry, the University of Paris-Est, the CNRS, GENOPOLE-Évry, the Institute for Complex Systems in Paris-Ile de France, the ANR white project AutoChem and the French working group GDR GPL/LTP.

## Appendix: An **MGS** Implementation of the Runge–Kutta Methods

The Runge–Kutta methods is a famous and widely used family of integration scheme for solving problems of the form:

$$\frac{d\mathbf{X}(t)}{dt} = f(\mathbf{X}(t), t) \quad \mathbf{X}(0) = X_0 \quad (10.16)$$

They are based on the techniques developed by C. Runge and M.W. Kutta at the beginning of the twentieth century.

In the following, we propose an **MGS** implementation of the classical explicit fourth order Runge–Kutta method (RK4). This example can be extended to most of other Runge–Kutta methods. As the Euler method, the RK4 allows the computation of a sequence of vectors  $\mathbf{X}_n$  with a more complex scheme reducing errors of approximations. Starting from an initial state  $\mathbf{X}_0$  at time  $t_0$ , a step of the RK4 is given by

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \frac{\Delta t}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4) \quad t_{n+1} = t_n + \Delta t \quad (10.17)$$

with

$$\begin{aligned} \mathbf{k}_1 &= f(\mathbf{X}_n, t_n) & \mathbf{k}_2 &= f\left(\mathbf{X}_n + \frac{\Delta t \mathbf{k}_1}{2}, t_n + \frac{\Delta t}{2}\right) \\ \mathbf{k}_3 &= f\left(\mathbf{X}_n + \frac{\Delta t \mathbf{k}_2}{2}, t_n + \frac{\Delta t}{2}\right) & \mathbf{k}_4 &= f(\mathbf{X}_n + \Delta t \mathbf{k}_3, t_n + \Delta t) \end{aligned} \quad (10.18)$$

The **MGS** implementation can be divided into 3 steps:

1. The computation of expressions  $\mathbf{X}_n + c\mathbf{k}$  where  $c$  is a coefficient (either  $\frac{\Delta t}{2}$  or  $\Delta t$ ) and  $\mathbf{k}$  takes the value  $\mathbf{k}_i$ .
2. The computation of  $\mathbf{X}_{n+1}$  knowing the  $\mathbf{k}_i$ .
3. The composition of the two first steps.

Steps (1) and (2) are straightforward and can be implemented as follows:

```
trans RK4a[c,k] = { x => x + c*k. (^x) }
trans RK4b[dt,k1,k2,k3,k4] = {
  x => x + (dt/6) ( k1. (^x) + 2*k2. (^x) + 2*k3. (^x)
    + k4. (^x) )
}
```

These transformations apply the required computations on each coordinate of  $\mathbf{X}$ . The final step consists in implementing the function RK4:

```
fun RK4[dt,f](X,t) = (
  let k1 = f(X,t) in
  let k2 = f(RK4a[c=dt/2,k=k1](X),t+dt/2) in
  let k3 = f(RK4a[c=dt/2,k=k2](X),t+dt/2) in
  let k4 = f(RK4a[c=dt,k=k3](X),t+dt) in
  (RK4b[dt=dt,k1=k1,k2=k2,k3=k3,k4=k4](X),t+dt)
)
```

## References

- F. Aurenhammer. Voronoi diagrams—A survey of a fundamental geometric data structure. *ACM Comput Surv*, 23(3):345–405, 1991
- F. Bailly and G. Longo. *Mathématiques et sciences de la nature*. Hermann, Paris, 2006
- J.-P. Banâtre and D. LeMetayer. Programming by multiset transformation. *Comm ACM*, 36(1):98, 1993
- J.-P. Banâtre, P. Fradet, and Y. Radenac. Generalised multisets for chemical programming. *Math Struct Comput Sci*, 16(4):557–580, 2006
- P. Barbier de Reuille, I. Bohn-Courseau, K. Ljung, H. Morin, N. Carraro, C. Godin, and J. Traas. Computer simulations reveal novel properties of the cell-cell signaling network at the shoot apex in Arabidopsis. *Proc Natl Acad Sc USA*, 103(5):1627–1632, 2006a
- K. De Cock, X. Zhang, M. F. Bugallo, and P. M. Djuric. Stochastic simulation and parameter estimation of first order chemical reactions. In *12th European Signal Processing Conference (EUSIPCO'04)*, 2003
- R. Durrett and S. Levin. The importance of being discrete (and spatial). *Theor Popul Biol*, 46(3):363–394, 1994
- M. Eden. A two-dimensional growth process. In *Proceedings of Fourth Berkeley Symposium on Mathematics, Statistics, and Probability, Vol. 4*, pages 223–239, 1961
- M. Eigen and P. Schuster. *The hypercycle: A principle of natural self-organization*. Springer, Berlin, 1979
- D. Endy. Foundations for engineering biology. *Nature*, 438:449–453, 2005
- E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems, LASL Report LA-1940 (5). Technical report, 1965. Reprinted in the collected work of E. Fermi, Vol. 2, pp. 977–988, 1965
- M. Fisher, G. Malcolm, and R. Paton. Spatio-logical processes in intracellular signalling. *BioSystems*, 55:83–92, 2000
- W. Fontana. Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Proceedings of the Workshop on Artificial Life (ALIFE'90)*, Vol. 5, pages 159–210, 1992
- W. Fontana and L. W. Buss. “The arrival of the fittest”: Toward a theory of biological organization. *Bull Math Biol*, 1994
- J.-L. Giavitto. Topological collections, transformations and their application to the modeling and the simulation of dynamical systems. In *Rewriting Techniques and Applications (RTA'03)*, LNCS 2706, pages 208–233, 2003
- J.-L. Giavitto and O. Michel. Declarative definition of group indexed data structures and approximation of their domains. In *Proceedings of the 3rd ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming (PPDP'01)*, pages 150–161, 2001
- J.-L. Giavitto and O. Michel. Data structure as topological spaces. In *Proceedings of the 3rd International Conference on Unconventional Models of Computation UMC'02*, Vol. LNCS 2509, pages 137–150, 2002a
- J.-L. Giavitto and O. Michel. Pattern-matching and rewriting rules for group indexed data structures. *ACM SIGPLAN Not*, 37(12):76–87, 2002b
- J.-L. Giavitto and O. Michel. The topological structures of membrane computing. *Fundam Inform*, 49:107–129, 2002c
- J.-L. Giavitto and O. Michel. Modeling the topological organization of cellular processes. *BioSystems*, 70(2):149–163, 2003
- J.-L. Giavitto and A. Spicher. Simulation of self-assembly processes using abstract reduction systems. In *Systems self-assembly: Sultidisciplinary snapshots*, pages 199–223. Elsevier, 2008a
- J.-L. Giavitto and A. Spicher. Topological rewriting and the geometrization of programming. *Physica D*, 237:1302–1314, 2008b
- J.-L. Giavitto, C. Godin, O. Michel, and P. Prusinkiewicz. Modelling and simulation of biological processes in the context of genomics. In *Computational Models for Integrative and Developmental Biology*, 2002a

- J.-L. Giavitto, O. Michel, and J. Cohen. Pattern-matching and rewriting rules for group indexed data structures. In *ACM Sigplan Workshop (RULE'02)*, pages 55–66, 2002b
- J.-L. Giavitto, G. Malcolm, and O. Michel. Rewriting systems and the modelling of biological systems. *Comp Funct Genomics*, 5:95–99, 2004
- M. C. Gibson, A. B. Patel, R. Nagpal, and N. Perrimon. The emergence of geometric order in proliferating metazoan epithelia. *Nature*, 442:1038–1041, 2006
- D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*, 81(25):2340–2361, 1977
- J. Greenberg and S. Hastings. Spatial patterns for discrete models of diffusion in excitable media. *SIAM J Appl Math*, pages 515–523, 1978
- Y. Itkis. *Control systems of variable structure*. Wiley, New York, 1976
- P. Jansson and J. Jeuring. PolyP – A polytypic programming language extension. In *Principles of programming languages*, pages 470–482. ACM, 1997
- T. Knight. Idempotent vector design for standard assembly of biobricks, 2006. MIT Synthetic Biology Working Group
- A. Lindenmayer. Mathematical models for cellular interaction in development, Parts I and II. *J Theor Biol*, 18:280–315, 1968a
- P. L. Luisi. Autopoiesis: A review and a reappraisal. *Naturwissenschaften*, 90:49–59, 2003
- J. Lynch. Logical characterization of individual-based models. In *23rd Annual IEEE Symposium on Logic in Computer Science (LICS'08)*, volume n, pages 379–390, 2008
- H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269(5224):650, 1995
- O. Michel, A. Spicher, and J.-L. Giavitto. Rule-based programming for integrative biological modeling – Application to the modeling of the  $\lambda$  phage genetic switch. *Nat Comput*, 8(4):865–889, 2009
- J. Munkres. *Elements of algebraic topology*. Addison-Wesley, Reading, MA, 1984
- G. Păun. From cells to computers: Computing with membranes (P systems). *Biosystems*, 59(3):139–158, 2001
- N.M. Shnerb, Y. Louzoun, E. Bettelheim, and S. Solomon. The importance of being discrete: Life always wins on the surface. *PNAS*, 97(19):10322–10324, 2000
- J. Smith. *Shaping life: Genes, embryos and evolution*. Yale University Press, New Haven, 1999
- A. Spicher and O. Michel. Using rewriting techniques in the simulation of dynamical systems: Application to the modeling of sperm crawling. In *Fifth International Conference on Computational Science (ICCS'05), Part I*, Vol. 3514 of LNCS, pages 820–827, 2005
- A. Spicher and O. Michel. Declarative modeling of a neurulation-like process. *BioSystems*, 87(2–3):281–288, 2007.
- A. Spicher, N. Fats, and O. Simonin. From reactive multi-agents models to cellular automata. In *International Conference on Agents and Artificial Intelligence*, pages 422–429, 2009
- A.M. Turing. The chemical basis of morphogenesis, Series B: Biological Sciences. *Phil Trans R Soc Lond*, 237:37–72, 1952
- G. Turk. Generating textures for arbitrary surfaces using reaction-diffusion. In T.W. Sederberg, editor, *Computer Graphics (SIGGRAPH '91 Proceedings)*, pages 289–298, 1991
- F.J. Varela, H.R. Maturana, and R. Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, 5:187–196, 1974
- G. Von Dassow, E. Meir, E. M. Munro, and G. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000
- J. Von Neumann. *Theory of self-reproducing automata*. University of Illinois Press, Urbana, 1966
- S. Wolfram. *Theory and applications of cellular automata*. World Scientific Publication, Singapore, 1986
- M. Wooldridge and M. Wooldridge. *Introduction to multiagent systems*. Wiley, New York, 2001
- X. Zhang, K. De Cock, M.F. Bugallo, and P.M. Djuric. Stochastic simulation and parameter estimation of enzyme reaction models. In *IEEE Workshop on Statistical Signal Processing*, 2003

# Glossary

**Antibiotic resistance** The ability of a microorganism to resist the effects of an antibiotic. Examples of mechanisms of antibiotic resistance include the synthesis of antibiotic-degrading enzymes (e.g.,  $\beta$ -lactamase), and modifications to drug targets such as the penicillin-binding proteins (PBPs) in bacterial cell membranes.

**BAIT** Bacteria-Antibiotic Interaction Tool (BAIT) is a precursor of Micro-Gen, which implements a simple model of bacterial growth and interactions with antibiotic molecules in a discrete, 2D environment.

**Bibliome** A collection composed of the primary and review literature, in addition to textbooks on a particular topic.

**Column space** For a matrix  $\mathbf{S}$  and vector  $\mathbf{v}$ , any vector that satisfies the relationship  $\mathbf{S} \times \mathbf{v} = b$ . When  $b = dx/dt$ , the time derivative of the concentration vector, then the column space describes metabolite dynamics.

**Delay differential equations with discrete time delays** Ordinary differential equations where the derivatives of the state variables at time  $t$  depend on the state variables at time  $t$  and possibly at a finite number of discrete times less than  $t$ . Such equations are functional differential equations of retarded type.

**Dynamic instability** The existence of two phases of growing and shrinking of microtubules in a population with rare transitions between these phases is referred to as dynamic instability.

**Dynamical system with a dynamical structure** A kind of dynamical system that requires its phase space, used to characterize the structure of the state of the system at each time step, to be computed jointly with the running state of the system.

**Endocrine hormone** A chemical substance produced by a ductless gland of the body, secreted into the blood stream and transported via the blood to affect other bodily organs having cell receptors specific to that hormone. For example, the pituitary gland secretes follicle stimulating hormone which promotes the growth of follicles within the ovaries.

**Global regulator** A transcription factor or another regulatory protein capable of controlling the expression of a large number of others genes.

**Genetic regulatory network** A network of genes, RNAs, proteins, metabolites, and their mutual regulatory interactions.

**Genome-scale** The characterization of a of biological function and components on spanning the genome of the respective organism, i.e., incorporation/consideration of all known associated components encoded in the organisms genome.

**Hill function** In biochemistry, the binding of a ligand to a macromolecule is referred to as *cooperative binding*. The Hill function (or Hill equation) is used to describe this effect. It is defined as  $y = K[x]^h / (1 + K[x]^h)$ , where  $y$ , the fractional saturation, is the fraction of the total number of binding sites occupied by the ligand,  $[x]$  is the free (unbound) ligand concentration,  $K$  is a constant, and  $h$  is the Hill coefficient.

**Integrative spatial systems biology** An emergent field in systems biology that deals with the necessary integration of spatial properties into integrative biology.

**Jacobian matrix** The Jacobian matrix refers to a matrix of all first-order partial derivatives of a vector-valued function. The Jacobian matrix represents the best linear approximation to a differentiable function near a given point.

**(Left) null space** For a matrix  $\mathbf{S}$  and vector  $\mathbf{v}$ , any vector that satisfies the relationship  $\mathbf{S}^T \times \mathbf{v} = 0$ , in which  $\mathbf{S}^T$  is the transpose of the matrix, is said to lie in the left null space of  $\mathbf{S}$ . This corresponds to conserved chemical moieties (i.e., compounds or chemical groups that are neither consumed nor produced in the network) in metabolic networks.

**Law of mass action** In chemistry, the law of mass action states that the rate of a chemical reaction is directly proportional to the molecular concentrations of the reacting substances. The law of mass action covers the equilibrium as well kinetic aspects (reaction rates) of chemical reactions.

**MGS** A domain-specific language aimed at the modeling and simulation of dynamical systems with a dynamical structure. MGS relies on the notions of local interactions, topological collections and their transformation in a declarative framework.

**Menstrual cycle** The reproductive cycle of a human female. The menstrual cycle describes the monthly changes which occur in a woman of reproductive age needed for the creation of an embryo. Under the control of endocrine hormones, follicles develop in the ovaries and then the dominant follicle releases its egg (ovulation) and becomes the corpus luteum which prepares the body for fertilization and implantation.

**MRSA** Methicillin-resistant *Staphylococcus aureus* (MRSA) is a multi-drug resistant form of *S. aureus* which was first isolated in 1961. Resistance conferred by expression of penicillin-binding protein 2a which has reduced binding to  $\beta$ -lactam antibiotics.

**Model checking** Algorithms that test the truth value of properties expressed in *temporal logic* on a *state transition graph*.

**Model reduction** The approximation of a model of a complex (non-linear) dynamical systems, with the aim of obtaining a simplified model that is easier to analyze but preserves essential properties of the original model.

**(Right) null space** For a matrix  $\mathbf{S}$  and vector  $\mathbf{v}$ , any vector that satisfies the relationship  $\mathbf{S} \times \mathbf{v} = 0$ , is said to lie in the right null space of  $\mathbf{S}$ . This corresponds to steady state flux distributions in metabolic networks.

**Row space** For a matrix  $\mathbf{S}$  and vector  $\mathbf{v}$ , any vector that satisfies the relationship  $\mathbf{S}^T \times \mathbf{v} = b$ . When  $b = dx/dt$ , the time derivative of the concentration vector, then the column space describes flux dynamics.

**Ordinary differential equation** In chemical kinetic theory, the interactions between species are commonly expressed using ordinary differential equations (ODEs). An ODE is a relation that contains *functions* of only one independent variable (typically  $t$ ), and one or more of its derivatives with respect to that variable. The order of an ODE is determined by the highest derivative it contains (for example, a first-order ODE involves only the first derivative of the function). The equation  $5x(t) + \dot{x}(t) = 17$  is an example of a first-order ODE involving the independent variable  $t$ , a function of this variable,  $x(t)$ , and a derivative of this function,  $\dot{x}(t)$ . Since a derivative specifies a rate of change, such an equation states how a function changes but does not specify the function itself. Given sufficient initial conditions, various methods are available to determine the unknown function. The difference between ordinary differential equations and partial differential equations is that partial differential equations involve partial derivatives of several variables.

**Partial differential equation** Is similar to an *ordinary differential equation* except that it involves functions with more than one independent variable.

**Piecewise-linear model** A set of linear differential equation models, each of which is defined on a part of the state space.

**Ran** A Ras related GTPase that can be in a GTP (guanosine tri-phosphate) or GDP (guanosine di-phosphate) bound state. It is implicated in nuclear import of proteins and cell division.

**RanGAP** Ran GTPase activating protein. It accelerates the hydrolysis of GTP by Ran.

**RCC1** Regulator of chromatin condensation 1 (RCC1) is a nuclear protein. Also referred to as RanGEF (Ran guanine nucleotide exchange factor). It is imported to the nucleus and binds to chromatin.

**Search and capture** Microtubules radially nucleated from a centrosome search cellular space till they encounter the centromere of a chromosome and are captured. This process is a model of events that lead to successful chromosome-microtubule linkage, critical in cell division and maintenance of ploidy.

**Sensitivity analysis** An important tool to study the dependence of systems on their parameters. Sensitivity analysis helps to identify those parameters that have signif-

icant impact on the system output and capture the essential characteristics of the system. Sensitivity analysis is particularly useful for complex biological networks with a large number of variables and parameters.

**State transition graph** A directed graph representing the behavior of a dynamic system in computer science. The nodes of the graph correspond to the states of the system, whereas the edges account for transitions from one state to another. More specifically, in the qualitative modeling of genetic regulatory networks using piecewise-linear models, the nodes describe a qualitative state of a network, consisting of the region in the state space in which the system resides, and the signs of the derivatives of the concentration variables in this region.

**Temporal logic** A formalism for describing sequences of transitions between states of a system, where the notion of time-order is introduced through the use of temporal operators.

**Topological collection** A collection of entities structured by a neighborhood relationships and handled as a whole. The fundamental data structure of MGS for the representation of the state of a dynamical system.

**Transformations** A set of rules acting on a topological collection. Used in MGS for the specification of the evolution function of a dynamical system.

**Zero-order ultrasensitivity** The chemical reactions in which the product formation is ultrasensitive and independent of substrate concentration since the enzyme is working at saturation (i.e. it depends on the zeroth power of substrate concentration) are seen to display zero-order ultrasensitivity.

**$\beta$ -lactamase** Enzyme produced by bacterial cells which degrades  $\beta$ -lactam antibiotic molecules by cleaving their  $\beta$ -lactam ring structure.

# Index

## A

ADME, 59

Analysis  
tool

Genetic Network Analyzer, 119

Antibiotic resistance, 132, 133, 138, 141, 147

## B

Bifurcation, 52

Bilinear interaction, 23, 34

Biological network reconstruction, 85, 89

Boolean network, vii, viii, 85, 102, 104, 112,  
125, 135

## C

CellDesigner, 78

Cellular automaton, 135

Contraceptives, 41

COPASI, 78

## D

Drug resistance, 75

## E

E. coli, 112

Estradiol, 40

## F

Flux balance analysis, viii, 20, 21, 89

Flux-concentration duality, 24, 26

Folate, viii, 27

## G

Genetic regulatory network, 112, 125

Global regulators of transcription, 112–114,  
119, 120, 123

Gradient matrix, 23–26, 28, 34–36

## H

Hill function, 44, 46, 53, 66, 72, 117

## I

Inhibin, 40, 45, 46, 51, 56

Interaction

antibiotic molecules – cell surface, 131

antibiotics-bacteria, 132, 137, 142

bacterial, 136

cell-cell, 179, 224

drug-cancer, 78

drug-system, 84

drug-target, 84

host-pathogen, 95, 102–105

human-virus, 83

kinetic, 26

protein-protein, 2, 84, 88, 96, 98

spatial, 196, 202, 227

Interactome, 96, 97, 101, 103

## J

Jacobian matrix, 23, 25, 35, 36

## K

Kinetics

Hill, 114

Michaelis-Menten, 114

## M

Mass action, 136, 221

kinetics, 28, 66

law, 221

Menstrual cycle, [viii](#), [39](#), [41](#), [45](#), [51](#), [53](#)  
Methionine, [28](#), [30](#), [31](#)  
Michaelis-Menten  
    constant, [8](#)  
    equation, [62](#), [74](#)  
Microtubule, [ix](#), [76–78](#), [156](#), [158](#), [160](#), [161](#),  
    [165](#), [168](#)  
Model  
    agent-based, [132](#), [136](#), [137](#), [147](#), [151](#)  
    kinetic, [19](#)  
    qualitative, [112](#)  
Model checking, [123](#), [124](#)  
Modeling  
    bottom-up, [64](#), [132](#)  
    multi-scale, [190](#)  
    tool  
        CellDesigner, [77](#)  
        COPASI, [77](#)  
MRSA, [132](#), [134](#), [142](#), [143](#)  
Mycobacterium tuberculosis, [83](#)  
  
**O**  
Ovulation, [39](#), [41](#), [42](#), [45](#), [53](#), [54](#), [56](#)

**P**  
Pituitary gland, [39](#), [42](#)

**S**  
Sensitivity analysis, [viii](#), [7](#), [9](#), [12](#), [14](#), [15](#), [50](#),  
    [77](#), [78](#), [193](#)  
Sensitivity coefficient, [52](#), [53](#)  
Simulation  
    bacteria-antibiotic interactions, [147](#)  
State transition graph, [117](#), [118](#), [120](#), [121](#), [123](#)  
Stoichiometric matrix, [20](#), [23](#), [26](#), [28](#), [30](#), [35](#),  
    [89](#), [93](#)

**T**  
Target identification, [2](#), [16](#), [84](#), [87](#), [89](#), [91](#), [93](#),  
    [94](#), [96–98](#), [106](#)  
Temporal logic, [121](#)  
Tuberculosis, [83](#), [100](#)