Werner Dubitzky · Jennifer Southgate
Hendrik Fuß *Editors*

Understanding the Dynamics of Biological Systems

Lessons Learned from Integrative Systems Biology

Springer

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

Springer

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

Hendrik Fuß
Nano Systems Biology Research Group
Biomedical Sciences Research Institute
University of Ulster
Coleraine BT52 1SA
United Kingdom
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

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)

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

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; cathale@iiserpune.ac.in

Valentina Baldazzi INRIA Grenoble – Rhône-Alpes, France, Valentina.Baldazzi@inria.fr

Nagasuma Chandra Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India, nchandra@serc.iisc.ernet.in

Hidde de Jong INRIA Grenoble – Rhône-Alpes, France, Hidde.de-Jong@inria.fr

Johannes Geiselmann Université Joseph Fourier, Grenoble, France, 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

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

Pedro T. Monteiro INRIA Grenoble – Rhône-Alpes, France
and
IST/INESC-ID, 9 Rua Alves Redol, 1000-029 Lisbon, Portugal, 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

Michel Page INRIA Grenoble – Rhône-Alpes, France, 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

R. Drew Pasteur Department of Mathematics and Computer Science, The College of Wooster, 1189 Beall Ave., Wooster, OH 44691, USA, 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

Karthik Raman Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India, k.raman@bioc.unizh.ch

Delphine Ropers INRIA Grenoble – Rhône-Alpes, France, 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

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

James F. Selgrade Department of Mathematics and Biomathematics Program, North Carolina State University, Raleigh, NC 27695-8205, USA, 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

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

Ray Walshe Centre for Scientific Computing and Complex Systems Modeling, School of Computing, Dublin City University, Dublin 9, Ireland, 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

W. Dubitzky et al. (eds.), *Understanding the Dynamics of Biological Systems: Lessons Learned from Integrative Systems Biology*, DOI 10.1007/978-1-4419-7964-3_1,
© Springer Science+Business Media, LLC 2011

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.

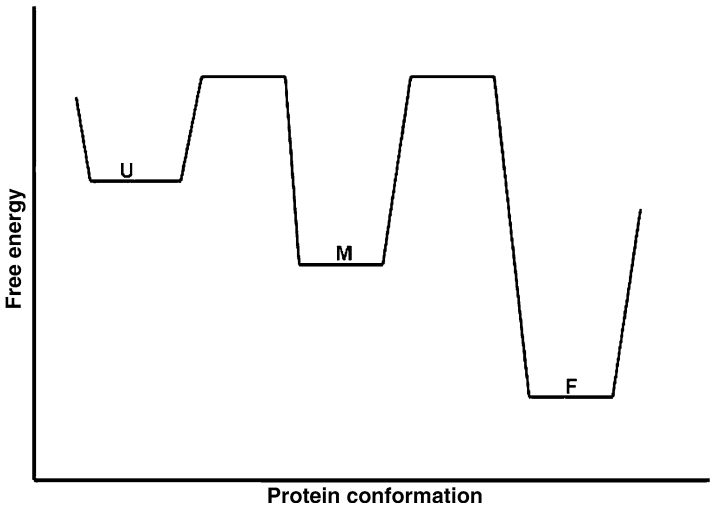

Fig. 1.1 *Different protein conformations have different energies.* While the goal is to reach the lowest energy as a properly folded protein (F), some misfolded proteins (M) are located in energy minima. Unfolded protein is denoted by U

A variety of diseases are the linked to protein misfolding. For example, disruption of proinsulin folding in β-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

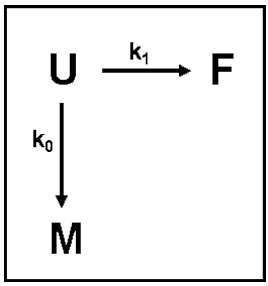

Fig. 1.2 *Schematic of protein folding without quality control.* The three protein conformations are represented as follows: unfolded (U), folded (F), and misfolded (M). Folding and misfolding reaction velocities are first-order with rate constants k_1 and k_0, respectively. We assume that folding and 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 ($u + m + f = constant$). 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.

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -(k_1 + k_0)\, u \tag{1.1}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = k_0\, u \tag{1.2}$$

$$\frac{\mathrm{d}\,f}{\mathrm{d}t} = k_1\, u. \tag{1.3}$$

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 ($\mathrm{u}(0) = u_0$, $\mathrm{m}(0) = 0$, and $\mathrm{f}(0) = 0$), we find the analytical solution of our system to be:

$$u(t) = u_0\, \mathrm{e}^{-(\mathrm{k}_0+\mathrm{k}_1)\,\mathrm{t}} \tag{1.4}$$

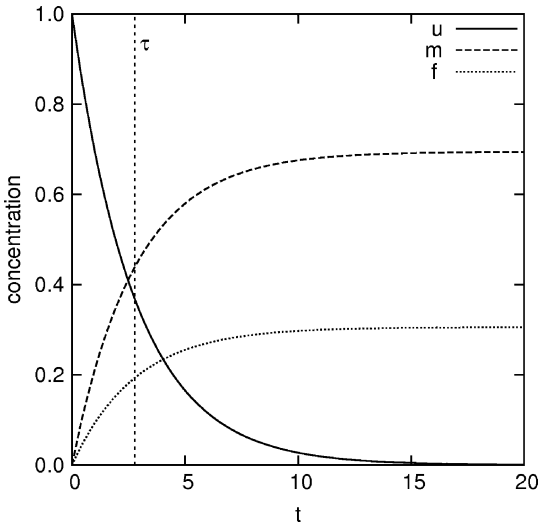

Fig. 1.3 *Time course of the three protein conformation concentrations in absence of quality control.* We begin with a basal unfolded protein (u) concentration, u_0, of 1 μM. Misfolded protein concentration, m, reaches a maximum of $\frac{k_0}{k_0+k_1}u_0$ and folded protein concentration, f, reaches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

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

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

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,

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

$$f_{\max} = \frac{k_1}{k_0 + k_1}u_0. \tag{1.8}$$

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):

$$\text{timescale of } x(t) \approx \frac{x_{\max} - x_{\min}}{\left|\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{\max}}. \tag{1.9}$$

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

$$\tau = \frac{1}{k_0 + k_1}. \tag{1.10}$$

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 β-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

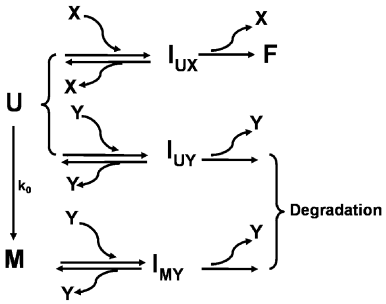

Fig. 1.4 Schematic of protein folding with quality control. U is unfolded protein, F is folded protein, and M is misfolded protein. X is the enzyme representative of the folding machinery. Y is the enzyme representative of the degradation machinery. The enzyme–substrate complex intermediate for each 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:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -k_0\, u - \frac{V_{\mathrm{f}}\, u}{K_{\mathrm{f}} + u} - \frac{V_{\mathrm{u}}\, u}{K_{\mathrm{u}}\left(1 + \frac{m}{K_{\mathrm{m}}}\right) + u} \tag{1.11}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = k_0\, u - \frac{V_{\mathrm{m}}\, m}{K_{\mathrm{m}}\left(1 + \frac{u}{K_{\mathrm{u}}}\right) + m} \tag{1.12}$$

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{V_{\mathrm{f}}\, u}{K_{\mathrm{f}} + u}. \tag{1.13}$$

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: $(u^\star, m^\star) = (0, 0)$. 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.

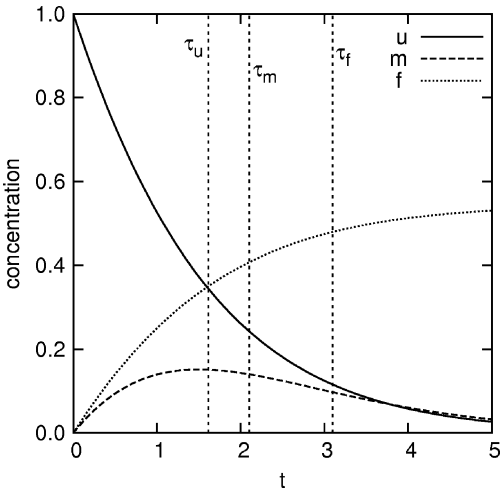

Fig. 1.5 *The time course of the unfolded protein (U), misfolded protein (M) and folded protein (F) concentrations under quality control.* The timescales for unfolded protein depletion (τ_u), misfolded protein depletion (τ_m), and folded protein production (τ_f) are denoted by the *vertical lines*. The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

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:

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

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

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

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