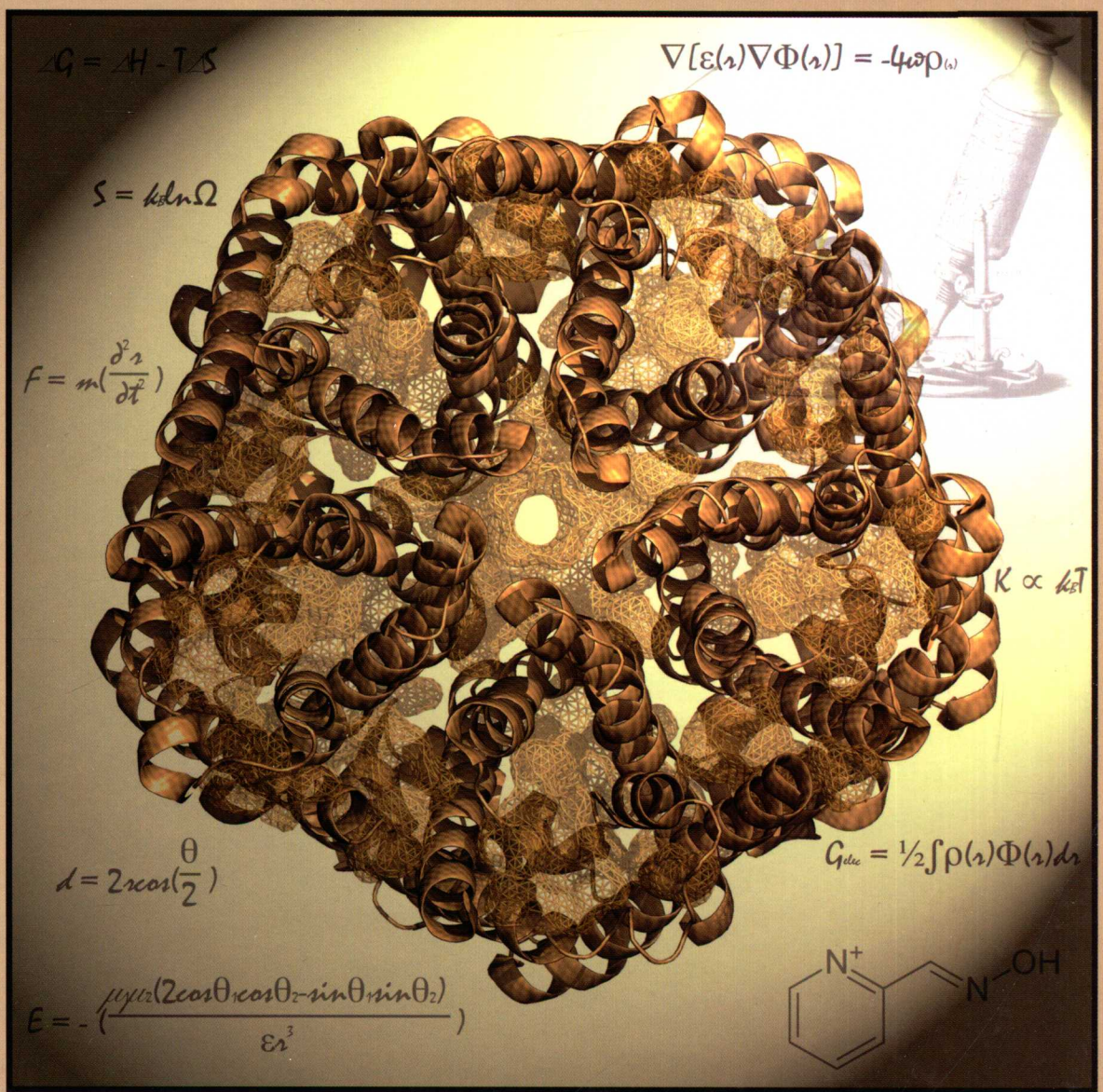
Introduction to Proteins

Structure, Function, and Motion



Amit Kessel and Nir Ben-Tal

CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Chapman & Hall/CRC Mathematical and Computational Biology Series

INTRODUCTION TO PROTEINS

STRUCTURE, FUNCTION, AND MOTION

AMIT KESSEL AND NIR BEN-TAL

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in India by Replika Press Pvt. Ltd.
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-1071-2 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Introduction to Proteins

Structure, Function, and Motion

CHAPMAN & HALL/CRC

Mathematical and Computational Biology Series

Aims and scope:

This series aims to capture new developments and summarize what is known over the entire spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical, and computational methods into biology by publishing a broad range of textbooks, reference works, and handbooks. The titles included in the series are meant to appeal to students, researchers, and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Series Editors

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
4th, Floor, Albert House
1-4 Singer Street
London EC2A 4BQ
UK

Published Titles

Algorithms in Bioinformatics: A Practical Introduction
Wing-Kin Sung

Bioinformatics: A Practical Approach
Shui Qing Ye

Biological Computation
Ehud Lamm and Ron Unger

Biological Sequence Analysis Using the SeqAn C++ Library
Andreas Gogol-Döring and Knut Reinert

Cancer Modelling and Simulation
Luigi Preziosi

Cancer Systems Biology
Edwin Wang

Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling
Arnaud Chauvière, Luigi Preziosi, and Claude Verdier

Clustering in Bioinformatics and Drug Discovery
John D. MacCuish and Norah E. MacCuish

Combinatorial Pattern Matching Algorithms in Computational Biology Using Perl and R
Gabriel Valiente

Computational Biology: A Statistical Mechanics Perspective
Ralf Blossey

Computational Hydrodynamics of Capsules and Biological Cells
C. Pozrikidis

Computational Neuroscience: A Comprehensive Approach
Jianfeng Feng

Data Analysis Tools for DNA Microarrays
Sorin Draghici

Differential Equations and Mathematical Biology, Second Edition
D.S. Jones, M.J. Plank, and B.D. Sleeman

Engineering Genetic Circuits
Chris J. Myers

Exactly Solvable Models of Biological Invasion
Sergei V. Petrovskii and Bai-Lian Li

Gene Expression Studies Using Affymetrix Microarrays
Hinrich Göhlmann and Willem Talloen

Glycome Informatics: Methods and Applications
Kiyoko F. Aoki-Kinoshita

Handbook of Hidden Markov Models in Bioinformatics
Martin Gollery

Introduction to Bioinformatics
Anna Tramontano

Introduction to Computational Proteomics
Golan Yona

Introduction to Proteins: Structure, Function, and Motion
Amit Kessel and Nir Ben-Tal

An Introduction to Systems Biology: Design Principles of Biological Circuits
Uri Alon

Kinetic Modelling in Systems Biology
Oleg Demin and Igor Goryanin

Knowledge Discovery in Proteomics
Igor Jurisica and Dennis Wigle

Meta-analysis and Combining Information in Genetics and Genomics
Rudy Guerra and Darlene R. Goldstein

Methods in Medical Informatics: Fundamentals of Healthcare Programming in Perl, Python, and Ruby
Jules J. Berman

Modeling and Simulation of Capsules and Biological Cells
C. Pozrikidis

Niche Modeling: Predictions from Statistical Distributions
David Stockwell

Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems
Qiang Cui and Ivet Bahar

Optimal Control Applied to Biological Models
Suzanne Lenhart and John T. Workman

Pattern Discovery in Bioinformatics: Theory & Algorithms
Laxmi Parida

Python for Bioinformatics
Sebastian Bassi

Spatial Ecology
Stephen Cantrell, Chris Cosner, and Shigui Ruan

Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation
Horst Malchow, Sergei V. Petrovskii, and Ezio Venturino

Stochastic Modelling for Systems Biology
Darren J. Wilkinson

Structural Bioinformatics: An Algorithmic Approach
Forbes J. Burkowski

The Ten Most Wanted Solutions in Protein Bioinformatics
Anna Tramontano

Preface

Proteins are highly complex molecules that are actively involved in the most basic and important aspects of life. These include metabolism, movement, defense, cellular communication, and molecular recognition. Accordingly, protein science is at the very center of biological research and is applied to areas such as medicine, agriculture, biotechnology, and even unconventional warfare.

In the last few decades, with the development of accurate and sophisticated means of molecular structure determination, it has become clear that the functions of macromolecules in general and of proteins in particular result directly from their structures and structural dynamics. In addition, it has been realized that true understanding of these aspects requires both qualitative and quantitative characterization of the dominant physical forces acting on proteins at the atomic level. This understanding has prompted a new field in biological sciences, termed "Structural Biophysics."

This book aims to provide the reader with a detailed description of protein structure and dynamics, combined with an in-depth discussion of the relationship between both these aspects and protein function. Adopting the structural–biophysical approach, we discuss these topics in relation to molecular interactions and thermodynamic changes that transpire in this highly complex system. There are several types of textbooks describing protein structure and function. Biochemistry textbooks emphasize the functional aspect of proteins and provide a rather general description of structure and structure–function relationship (SFR). Structural Biology textbooks provide an extensive description of protein structure and also refer to SFR with varying degrees of detail. However, energy-related aspects are often avoided. Molecular biophysics textbooks focus on molecular interactions and thermodynamic aspects of protein structure, but tend to lack detailed description of structural and dynamic aspects, as well as SFR. This book refers to all the aforementioned aspects and attempts to provide a unified view. Our energy-oriented approach is manifested throughout the book, whether we discuss structure, dynamics or specific functions of proteins. An extensive discussion of the energetics of protein structure is also given in a chapter dedicated to this topic.

Clarification of SFR in proteins is the ultimate goal of any book about proteins. Most textbooks describe it *via* specific examples or protein types. This gives the readers a wide view on protein activity, but general conclusions are often missing. Here we attempt to provide the readers with the very principles of protein action (as we understand it), in the form of guidelines, when possible. This is done throughout the book, but particularly in the

chapter describing protein–ligand interactions. We chose to elaborate on this issue since, in our view, it is the very basis of all protein functions.

The central dogma of structural biology is the dependence of function on structure. Yet, some proteins, termed "intrinsically unstructured proteins" (IUPs), are inherently devoid of regular three-dimensional structure, and still have numerous functions. IUPs have been extensively studied in the last few years, yet are not mentioned in most textbooks. We dedicated a chapter of the book to these proteins in order to provide a more complete and realistic view of the proteome, as well as to explore the full repertoire of protein functions.

Much of the knowledge on protein structure and function relationship became possible only after technologies for the determination or prediction of three-dimensional protein structure emerged. Accordingly, we provide a concise description of the main experimental and computational methods used today for studying protein structure and dynamics. In this respect, we mention various Internet-based resources, such as databases and algorithms, which are widely used and fully accessible to the reader. As mentioned above, protein science is not only of academic interest. Indeed, it has been applied in various industrial, medical, and agricultural fields. In our book we discuss two of these applications: the industrial use of enzymes and protein engineering, and the rational design of protein-targeting pharmaceutical drugs. Based on the above, we believe that the book will be of interest to both students and scientists in protein-related fields.

Proteins: Structure, Function, and Motion is intended for various audiences. First, it can be used by undergraduate or graduate students of biochemistry, structural biology, computational biophysics, bioinformatics, and biotechnology as an introduction to protein structure. In that sense, it may serve as a standalone textbook for basic-to-intermediate level courses in structural biology. For such purposes, we have included exercises on theory and practice. Second, we expect the parts of the book discussing in detail energetic, dynamic, and evolutionary aspects of proteins to be of special interest to post-graduate scientists and to industry professionals. To make it easier for these two groups of readers to find their texts of interest, we have, in some cases, separated the basic material from the more advanced material, by putting the latter in boxes. Finally, the book refers to many everyday issues related to proteins and enzymes, such as medical disorders, drugs, toxins, chemical warfare, and animal behavior. We hope these topics will create interest among non-professional science enthusiasts as well.

The following is a general outline of the book:

Chapter 1, "Introduction", includes three parts. The first provides an overview of proteins' main functions and their importance to various fields, e.g., medicine and the drug industry. The second explains the central "structure–dynamics–function" paradigm in proteins, thus providing the general rationale for the book. The third part describes the non-covalent forces acting on macromolecules, an overview that is needed for understanding the notions presented later on in the book. Finally, the general layout of the book is presented.

Chapter 2, "Protein Structure", describes in detail the different levels of protein structure. The physico-chemical properties of amino acids are described at length. The description of secondary, tertiary, and quaternary structure that follows emphasizes the structural principles achieved by the observed architecture. Other factors affecting both protein structure and function, i.e., non-natural amino acids, enzymatic cofactors, prosthetic groups, and post-translational modifications, are also described, with emphasis on SFR. All of these topics are exemplified using specific proteins. For instance, *protein kinase A* (*PKA*), a central enzyme in cellular communication, is used to demonstrate some of the main advantages of quaternary structure. Pyruvate dehydrogenase, a large enzyme complex involved in carbohydrate metabolism, is used to demonstrate the role of cofactors and prosthetic groups in protein function.

Chapter 3, "Methods of Structure Determination and Prediction", describes the main methods used today for structure determination and their applications. First, methods based on particle/wave diffraction or scattering are described. These include *X-ray crystallography*, *neutron scattering*, and *electron scattering*. Then, spectroscopic methods, including *nuclear magnetic resonance* (*NMR*) *spectroscopy*, *electron paramagnetic resonance* (*EPR*) *spectroscopy*, and *circular dichroism* (*CD*), are presented. This is followed by a description of computational methods for predicting protein structure. It refers to the two main approaches in this field. The first, "physical" approach relies on a mathematical description of the physical forces acting on the protein's atoms. In this respect we describe well-known methods, such as *molecular dynamics* and *simulated annealing*. The second, "comparative" approach, the most prominent of which is *homology modeling*, relies on sequence comparisons and statistical data. We dedicate a great deal of this discussion to analyzing the advantages and disadvantages of each method, and the cases to which it is best applied. Finally, the current tools for comparing the different methods and evaluating their efficiency are presented.

Chapter 4, "Energetics and Protein Stability", discusses the thermodynamic aspects of protein structure. It begins with an overview of the basic thermodynamic variables, the ways they can be measured or calculated, and their interpretation in molecular systems. In discussing the latter, we refer to biological processes that can be characterized using thermodynamic variables. This includes metabolic processes, protein folding, and protein–ligand interactions. The second section of the chapter discusses the main physical forces in the system with respect to their influence on protein structure. In the third and fourth sections we examine two cases in which the theoretical principles discussed are applied. The first is the adaptation of unicellular organisms to extreme environments, and the second is the use of protein engineering to enhance the industrial uses of enzymes.

Chapter 5, "Protein Structural Dynamics", widens the structure–function paradigm by incorporating structural dynamics. Two aspects of protein dynamics are discussed: protein folding and folded (native) state dynamics. For the first, we present

the current views on how proteins acquire their three-dimensional structures. This area has been extensively studied, and we present the main conclusions. In addition, we discuss some well-known pathologies involving protein misfolding, such as *cystic fibrosis*, *Parkinson's disease*, *and mad cow disease*. The part that follows discusses changes that occur in the protein's native structure over time and illustrates their functional importance on different levels. In this respect, we elaborate on allostery, as a cellular approach for regulating proteins' function by manipulating their dynamic properties. We discuss different models and mechanisms of allostery and use specific proteins to demonstrate them. For example, we mention studies on the medically important enzyme *dihydrofolate reductase* (*DHFR*), in which long-distance allosteric effects were discovered. The oxygen-carrying protein hemoglobin is used to demonstrate in detail multi-level changes in protein dynamics, induced by allosteric regulators.

Chapter 6, "Non-Globular Proteins", focuses on two groups of proteins that seem to deviate from the "globular" behavior presented in the previous chapters. The first group includes proteins that play relatively simple roles inside and outside cells, forming large fibrous structures. We discuss some well-studied examples, such as *collagen*, the principal protein of connective tissues, and *keratin*, a protein that provides toughness to horns, nails, and claws. The second group of proteins includes those that are characterized by the absence of regular tertiary structure. These interesting proteins have many different functions that do not seem to require a permanent structure. As in Chapter 2, we discuss the principal properties of fibrous and unstructured proteins, with emphasis on SFR.

Chapter 7, "Membrane Proteins", focuses on a subtype of globular proteins that are located near or inside cellular membranes. These proteins constitute 20–30% of the human genome and play numerous roles in cellular physiology. Surrounded by a lipid environment, they are subjected to different forces than are water-soluble globular proteins, and therefore behave differently. The first part of the chapter overviews the structure, organization, and function of biological membranes. In this respect, it discusses membrane asymmetry and the variability of membrane composition (and hence, its properties) among different organisms. The second part analyzes membrane proteins, emphasizing common sequence- and structure-related themes, as well as folding energetics. The third part discusses the important issue of protein–membrane interactions, which has implications for both structure and function of membrane proteins. Finally, to illustrate SFR in membrane proteins, we focus on *GPCR*s, a group of receptors that serve as targets of most pharmacological drugs. We discuss in detail the β-*adrenergic receptor*, the structure of which has been determined recently in its active state. Membrane proteins are notoriously difficult to crystallize, and are therefore a desirable target for structure prediction. Throughout the chapter we mention key computational approaches developed for locating membrane proteins within genomes, and for predicting their topology and full three-dimensional structure.

Chapter 8, "Protein–Ligand Interactions", demonstrates SFR in proteins by addressing their most important ability, i.e., binding to other molecules. After a short overview of the functional aspects of this ability, we discuss past and present theories on binding, and its thermodynamic implications. We then analyze the binding on a molecular level, by focusing on the properties of protein-binding sites. One such property is the electrostatic potential, which we demonstrate using the example of *acetylcholine esterase* (*AChE*). AChE is a major enzyme responsible for the correct functioning of the nervous system, and is, therefore, also a major target for various nerve agents and toxins. It is extremely fast, in part thanks to the mechanism of "electrostatic steering," which it uses to draw its natural substrate into the catalytic site. The following part in the chapter demonstrates the principles discussed above by addressing the example of protein–protein interactions. Finally, we discuss the rational design of pharmaceutical drugs, which is one of the most practical applications of protein–ligand interactions.

In this book, we use numerous proteins as examples, demonstrating the various topics and principles discussed. Some proteins are mentioned in different contexts, to reflect the multiple ways in which proteins can be studied and analyzed. For example, *hemoglobin* is used to demonstrate quaternary structure, pathologies stemming from structure-altering mutations, and the role of dynamics in allosteric regulation. Another example is the cancer-related protein *ras*, used to demonstrate different types of post-translational modifications.

Questions for each chapter are located in the "download" section of this book's Web page on the CRC Press Web site (http://www.crcpress.com/product/isbn/9781439810712). Qualifying instructors can contact the publisher to obtain answers to these questions as well as slides.

Acknowledgments

The authors would like to thank the following individuals:

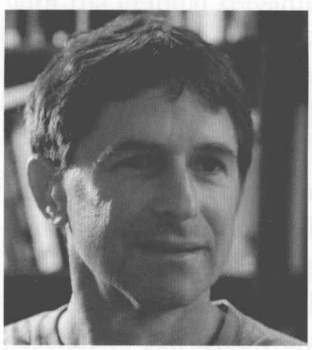
For helpful discussions: Boaz Shalem, MSc, Dr. Yfat Kessel-Kaufman, Dr. Avner Schlessinger, Dr. Shaul Shalem, Dr. Sarel Fleishman, Yosi Haim, MSc, and Matan Kalman, MSc.

For technical support: Varda Vexler, Maya Schushan, MSc, Iddo Better-Pocker and Daphna Meroz, MSc.

A.K. would also like to thank the following people for their ongoing personal support: Rachel and Nathan Stempler, Sara Kessel, Yfat and Eyal Kaufman, Boaz and Dafi Shalem.

Authors

Dr. Amit Kessel obtained his master's degree in experimental biochemistry at Tel-Aviv University studying the innate response of human blood cells to pathogenic bacteria. During his PhD studies he trained as a computational biophysicist, investigating the molecular basis of peptide–membrane interactions and the mechanisms of anti-bacterial peptides. In his post-doctoral research at Columbia University, Dr. Kessel continued studying proteins at the molecular level, focusing on various physicochemical aspects of protein–protein interactions. Today, he teaches protein biochemistry and biophysics at the Tel-Aviv-Yaffo Academic College, and has recently co-founded *Es-is Technologies Ltd.*, a company that designs biocatalysts for the pharmaceutical industry.



Professor Nir Ben-Tal obtained his bachelor's degree in biology, chemistry, and physics at the Hebrew University and his DSc in chemistry at Technion, Israel Institute of Technology. He carried out his post-doctoral training as a computational biophysicist at Columbia University and later joined the Department of Biochemistry and Molecular Biology at Tel-Aviv University, where he is currently a professor. His research includes various aspects in computational biology with a focus on structural bioinformatics. For example, his laboratory predicted the three-dimensional structures of a number of transmembrane proteins, thereby providing molecular insight into their mechanisms. His lab also develops the ConSurf Web-server for the detection of functional regions by mapping evolutionary data on protein structures.

Table of Contents