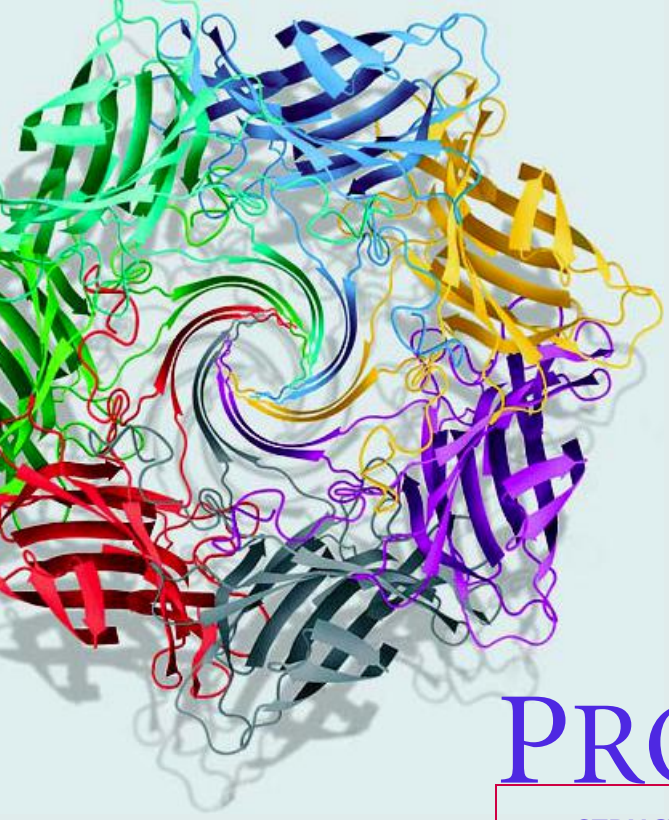
David Whitford

PROTEINS

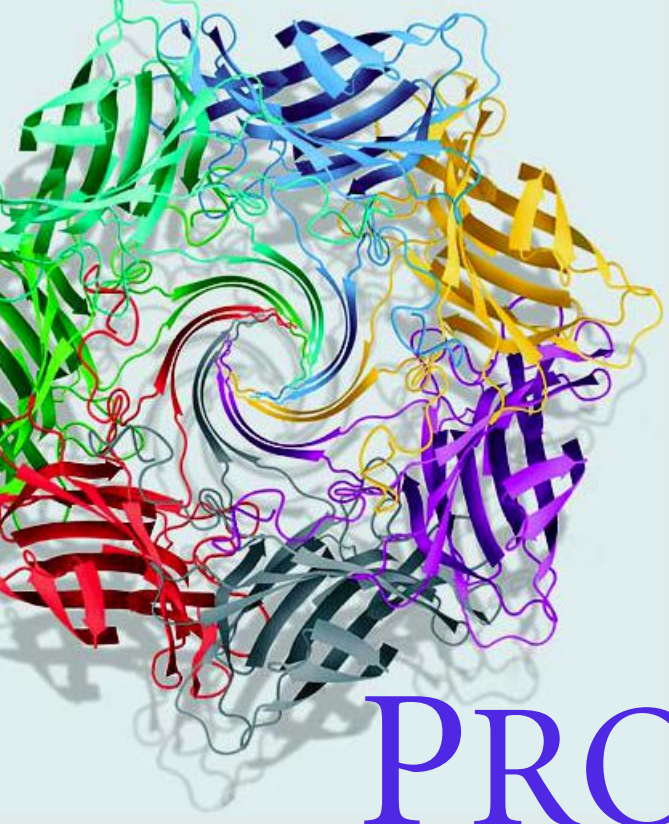
STRUCTURE AND FUNCTION

WILEY



PROTEINS

STRUCTURE AND FUNCTION



PROTEINS

STRUCTURE AND FUNCTION

David Whitford

John Wiley & Sons, Ltd

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-49893-9 HB
ISBN 0-471-49894-7 PB

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India
Printed and bound by Graphos SpA, Barcelona, Spain
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

For my parents,
Elizabeth and Percy Whitford,
to whom I owe everything

Contents

Preface

When I first started studying proteins as an undergraduate I encountered for the first time complex areas of biochemistry arising from the pioneering work of Pauling, Sumner, Kendrew, Perutz, Anfinsen, together with other scientific 'giants' too numerous to describe at length in this text. The area seemed complete. How wrong I was and how wrong an undergraduate's perception can be! The last 30 years have seen an explosion in the area of protein biochemistry so that my 1975 edition of *Biochemistry* by Albert Lehninger remains, perhaps, of historical interest only. The greatest change has occurred through the development of molecular biology where fragments of DNA are manipulated in ways previously unimagined. This has enabled DNA to be sequenced, cloned, manipulated and expressed in many different cells. As a result areas of recombinant DNA technology and protein engineering have evolved rapidly to become specialist disciplines in their own right. Almost any protein whose primary sequence is known can be produced in large quantity via the expression of cloned or synthetic genes in recombinant host cells. Not only is the method allowing scientists to study some proteins for the first time but the increased amount of protein derived from recombinant DNA technology is also allowing the application of new and continually advancing structural techniques. In this area X-ray crystallography has remained at the forefront for over 40 years as a method of determining protein structure but it is now joined by nuclear magnetic resonance (NMR) spectroscopy and more recently by cryoelectron microscopy whilst other methods such as circular dichroism, infrared and Raman spectroscopy, electron spin resonance spectroscopy, mass spectrometry and fluorescence provide more limited, yet often vital and complementary, structural data. In many instances these methods have become established techniques only in the last 20 years and are consequently absent in many of those familiar textbooks occupying the shelves of university libraries.

An even greater impact on biochemistry has occurred with the rapid development of cost-effective, powerful, desktop computers with performance equivalent to the previous generation of supercomputers. Many experimental techniques relied on the co-development of computer hardware but software has also played a vital role in protein biochemistry. We can now search databases comparing proteins at the level of DNA or amino acid sequences, building up patterns of homology and relationships that provide insight into origin and possible function. In addition we use computers routinely to calculate properties such as isoelectric point, number of hydrophobic residues or secondary structure – something that would have been extraordinarily tedious, time consuming and problematic 20 years ago. Computers have revolutionized all aspects of protein biochemistry and there is little doubt that their influence will continue to increase in the forthcoming decades. The new area of bioinformatics reflects these advances in computing.

In my attempt to construct an introductory yet extensive text on proteins I have, of necessity, been circumspect in my description of the subject area. I have often relied on qualitative rather than quantitative descriptions and I have attempted to minimise the introduction of unwieldy equations or formulae. This does not reflect my own interests in physical biochemistry because my research, I hope, was often quantitative. In some cases particularly the chapters on enzymes and physical methods the introduction of equations is unavoidable but also necessary to an initial description of the content of these chapters. I would be failing in my duty as an educator if I omitted some of these equations and I hope students will keep going at these 'difficult' points or failing that just omit them entirely

on first reading this book. However, in general I wish to introduce students to proteins by describing principles governing their structure and function and to avoid over-complication in this presentation through rigorous and quantitative treatment. This book is firmly intended to be a broad introductory text suitable for undergraduate and postgraduate study, perhaps after an initial exposure to the subject of protein biochemistry, whilst at the same time introducing specialist areas prior to future advanced study. I hope the following chapters will help to direct students to the amazing beauty and complexity of protein systems.

Target audience

The present text should be suitable for all introductory modes of biochemistry, molecular biology, chemistry, medicine and dentistry. In the UK this generally means the book is suitable for all undergraduates between years 1 and 3 and this book has stemmed from lectures given as parts of biochemistry courses to students of biochemistry, chemistry, medicine and dentistry in all 3 years. Where possible each chapter is structured to increase progressively in complexity. For purely introductory courses as would occur in years 1 or 2 it is sufficient to read only the first parts, or selected sections, of each chapter. More advanced courses may require thorough reading of each chapter together with consultation of the bibliography and secondly the list of references given at the end of the book.

The world wide web

In the last ten years the world wide web (WWW) has transformed information available to students. It provides a new and useful medium with which to deliver lecture notes and an exciting and new teaching resource for all. Consequently within this book URLs direct students to learning resources and a list of important addresses is included in the appendix. In an effort to exploit the power of the internet this book is associated with 'web-based' tutorials, problems and content and is accessed from the following URL http://www.wiley.com/go/whitfordproteins. These 'pages' are continually updated and point the interested reader towards new areas as they emerge. The Bibliography points interested readers towards further study material suitable for a first introduction to a subject whilst the list of references provides original sources for many areas covered in each of the twelve chapters.

For the problems included at the end of each chapter there are approximately 10 questions that aim to build on the subject matter discussed in the preceding text. Often the questions will increase in difficulty although this is not always the case. In this book I have limited the bibliography to broad reviews or accessible journal papers and I have deliberately restricted the number of 'high-powered' (difficult!) articles since I believe this organization is of greater use to students studying these subjects for the first time. To aid the learning process the web edition has multiple-choice questions for use as a formative assessment exercise. I should certainly like to hear of all mistakes or omissions encountered in this text and my hope is that educators and students will let me know via the e-mail address at the end of this section of any required corrections or additions.

Proteins are three-dimensional (3D) objects that are inadequately represented on book pages. Consequently many proteins are best viewed as molecular images using freely available software. Here, real-time manipulation of coordinate files is possible and will prove helpful to understanding aspects of structure and function. The importance of viewing, manipulating and even changing the representation of proteins to comprehending structure and function cannot be underestimated. Experience has suggested that the use of computers in this area can have a dramatic effect on student's understanding of protein structures. The ability to visualize in 3D conveys so much information – far more than any simple 2D picture in this book could ever hope to portray. Alongside many figures I have written the Protein DataBank files (e.g. PDB: 1HKO) used to produce diagrams. These files can be obtained from databases at several permanent sites based around the world such as http://www.rscb.org/pdb or one of the many 'mirrors' that exist (for example, in the UK this data is found at http://pdb.ccdc.cam.ac.uk). For students with Internet access each PDB file can be retrieved and manipulated independently to produce comparable images to those shown in the text. To explore these macromolecular images with reasonable efficiency does not require the latest 'all-powerful' desktop computer. A computer with a Pentium III (or later) based processor, a clock speed of 200 MHz or

greater, 32–64 MB RAM, hard disks of 10 GB, a graphics video card with at least 8 MB memory and a connection to the internet are sufficient to view and store a significant number of files together with representative images. Of course things are easier with a computer with a surfeit of memory (>256 MB) and a high 'clock' speed (>2 GHz) but it is not obligatory to see 'on-line' content or to manipulate molecular images. This book was started on a 700 MHz Pentium III based processor equipped with 256 MB RAM and 16 MB graphics card.

Organization of this book

This book will address the structure and function of proteins in 12 subsequent chapters each with a definitive theme. After an initial chapter describing why one would wish to study proteins and a brief historical background the second chapter deals with the 'building blocks' of proteins, namely the amino acids together with their respective chemical and physical properties. No attempt is made at any point to describe the metabolism connected with these amino acids and the reader should consult general textbooks for descriptions of the synthesis and degradation of amino acids. This is a major area in its own right and would have lengthened the present book too much. However, I would like to think that students will not avoid these areas because they remain an equally important subject that should be covered at some point within the undergraduate curriculum. Chapter 3 covers the assembly of amino acids into polypeptide chains and levels of organizational structure found within proteins. Almost all detailed knowledge of protein structure and function has arisen through studies of globular proteins but the presence of fibrous proteins with different structures and functional properties necessitated a separate chapter devoted to this area (Chapter 4). Within this class the best understood structures are those belonging to the collagen class of proteins, the keratins and the extended β sheet structures such as silk fibroin. The division between globular proteins and fibrous proteins was made at a time when the only properties one could compare readily were a protein's amino acid composition and hydrodynamic radius. It is now apparent that other proteins exist with properties intermediate between globular and fibrous proteins that do not lend themselves to simple classification. However, the 'old' schemes of identification retain their value and serve to emphasize differences in proteins.

Membrane proteins represent a third group with different composition and properties. Most of these proteins are poorly understood, but there have been spectacular successes from the initial low-resolution structure of bacteriorhodopsin to the highly defined structure of bacterial photosynthetic reaction centres. These advances paved the way towards structural studies of G proteins and G-protein coupled receptors, the respiratory complexes from aerobic bacteria and the structure of ATP synthetases.

Chapter 6 focuses both on experimental and computational methods of comparing proteins where *in silico* methods have become increasingly important as a vital tool to assist with modern protein biochemistry. Chapter 7 focuses on enzymes and by discussing basic reaction rate theories and kinetics the chapter leads to a discussion of enzyme-catalysed reactions. Enzymes catalyse reactions through a variety of mechanisms including acid–base catalysis, nucleophilic driven chemistry and transition state stabilization. These and other mechanisms are described along with the principles of regulation, active site chemistry and binding.

The involvement of proteins in the cell cycle, transcription, translation, sorting and degradation of proteins is described in Chapter 8. In 50 years we have progressed from elucidating the structure of DNA to uncovering how this information is converted into proteins. The chapter is based around the structure of two macromolecular systems: the ribosome devoted towards accurate and efficient synthesis and the proteasome designed to catalyse specific proteolysis. Chapter 9 deals with the methods of protein purification. Very often, biochemistry textbooks describe techniques without placing the technique in the correct context. As a result, in Chapter 9 I have attempted to describe equipment as well as techniques so that students may obtain a proper impression of this area.

Structural methods determine the topology or fold of proteins. With an elucidation of structure at atomic levels of resolution comes an understanding of biological function. Chapter 10 addresses this area by describing different techniques. X-ray crystallography remains at the forefront of research with new variations of the basic principle allowing faster determination of

structure at improved resolution. NMR methods yield structures of comparable resolution to crystallography for small soluble proteins. In ideal situations these methods provide complete structural determination of all heavy atoms but they are complemented by other spectroscopic methods such as absorbance and fluorescence methods, mass spectrometry and infrared spectroscopy. These techniques provide important ancillary information on tertiary structure such as the helical content of the protein, the proportion and environment of aromatic residues within a protein as well as secondary structure content.

Chapter 11 describes protein folding and stability – a subject that has generated intense research interest with the recognition that disease states arise from aberrant folding or stability. The mechanism of protein folding is illustrated by *in vitro* and *in vivo* studies. Whilst the broad concepts underlying protein folding were deduced from studies of 'model' proteins such as ribonuclease, analysis of cell folding pathways has highlighted specialised proteins, chaperones, with a critical function to the overall process. The GroES–GroEL complex is discussed to highlight the integrated process of synthesis and folding *in vivo*.

The final chapter builds on the preceding 11 chapters using a restricted set of well-studied proteins (case studies) with significant impact on molecular medicine. These proteins include haemoglobin, viral proteins, p53, prions and α_1-antitrypsin. Although still a young subject area this branch of protein science will expand in the next few years and will rely on the techniques, knowledge and principles elucidated in Chapters 1–11. The examples emphasize the impact of protein science and molecular medicine on the quality of human life.

Acknowledgements

I am indebted to all research students and post-docs who shared my laboratories at the Universities of London and Oxford during the last 15 years in many cases acting as 'test subjects' for teaching ideas. I should like to thank Drs Roger Hewson, Richard Newbold and Susan Manyusa whose comments throughout my research and teaching career were always valued. I would also like to thank individuals, too numerous to name, with whom I interacted at King's College London, Imperial College of Science, Technology and Medicine and the University of Oxford. In this context I should like to thank Dr John Russell, formerly of Imperial College London whose goodwill, humour and fantastic insight into the history of science, the scientific method and 'day to day' experimentation prevented absolute despair.

During preparation of this book many individuals read and contributed valuable comments to the manuscript's content, phrasing and ideas. In particular I wish to thank these unnamed and some times unknown individuals who read one or more of the chapters of this book. As is often said by most authors at this point despite their valuable contributions all of the remaining errors and deficiencies in the current text are my responsibility. In this context I could easily have spent more months attempting to perfect the current text. I am very aware that this text has deficiencies but I hope these defects will not detract from its value. In addition my wish to try other avenues, other roads not taken, dictates that this manuscript is completed without delay.

Writing and producing a textbook would not be possible without the support of a good publisher. I should like to thank all the staff at John Wiley & Sons, Chichester, UK. This exhaustive list includes particularly Andrew Slade as senior Publishing Editor who helped smooth the bumpy route towards production of this book, Lisa Tickner who first initiated events leading to commissioning this book, Rachel Ballard who supervised day to day business on this book, replacing every form I lost without complaint and monitoring tactfully and gently about possible completion dates, Robert Hambrook who translated my text and diagrams into a beautiful book, and the remainder of the production team of John Wiley and Sons. Together we inched our way towards the painfully slow production of this text, although the pace was entirely attributable to the author.

Lastly I must also thank Susan who tolerated the protracted completion of this book, reading chapters and offering support for this project throughout whilst coping with the arrival of Alexandra and Ethan effortlessly (unlike their father).

David Whitford
April 2004
david.whitford@ntlworld.com