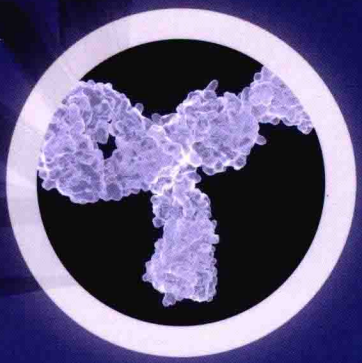
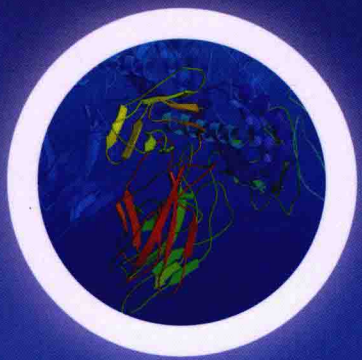
SECOND EDITION





Gary Walsh

Proteins

Biochemistry and Biotechnology

WILEY Blackwell

Proteins

Biochemistry and Biotechnology

Second Edition

Gary Walsh
Industrial Biochemistry Programme,
CES Department,
University of Limerick, Ireland

常州大学图书馆
藏书章

WILEY Blackwell

This edition first published 2014 © 2014 by John Wiley & Sons, Ltd

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

Editorial Offices
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Walsh, Gary (Biochemist), author.
Proteins : biochemistry and biotechnology / Gary Walsh. – 2e.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-66986-0 (cloth) – ISBN 978-0-470-66985-3 (pbk.)
I. Title.
[DNLM: 1. Proteins–chemistry. 2. Enzymes. 3. Industrial Microbiology. 4. Proteins–analysis. 5. Proteins–therapeutic use. QU 55]
TP248.65.P76
660.6'3–dc23

2013046817

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

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

Set in 10.5/12.5pt Minion by SPi Publisher Services, Pondicherry, India
Printed and bound in Singapore by Markono Print Media Pte Ltd

1 2014

Proteins

This book is dedicated to a most precious collection of proteins, my children Eithne, Shane and Alice.

此为试读，需要完整PDF请访问：www.ertongbook.com

Preface

This textbook aims to provide a comprehensive and up-to-date overview of proteins, both in terms of their biochemistry and applications. The first edition was published over a decade ago and in the intervening period this field has continued to rapidly evolve. The new edition retains the overall structure of the original one. Chapters 1–4 are largely concerned with basic biochemical principles. In these chapters issues relating to proteomics, protein sources, structure, engineering, purification and characterization are addressed. The remaining 10 chapters largely focus on the production of proteins and their applications in medicine, analysis and industry.

Despite the similarity in overall structure, the new edition has been extensively revised and updated to reflect recent progress in the area. Relative to the earlier edition there is greater emphasis on protein biochemistry, engineering and proteomics. The production of proteins via fermentation and animal cell culture are considered in new sections, which better balance the subsequent consideration of protein purification. The protein application chapters have been updated to reflect recent trends and developments. Thus, for example, recent bioprocess developments such as the use of disposable bioreactors are considered, there is greater relative emphasis on recombinant production systems and engineered products, therapeutic antibodies now are considered in a full dedicated chapter, and newer industrial applications such as the use of enzymes in biofuel generation are also included. The chapters considering protein applications have also been strengthened via the incorporation of numerous specific commercial product case studies.

The text caters mainly for advanced undergraduate and graduate students undertaking courses in applied biochemistry/biotechnology, but it should also be of value to students pursuing degrees in biochemistry, microbiology, or any branch of the biomedical sciences. Its scope also renders it of interest to those currently working in the biotechnology sector.

A sincere note of thanks is due to a number of people who have contributed to the successful completion of this project. Thank you to J.J. Tobin, Tewfik Soulimane and Jayne Murphy for useful scientific discussions and to Angela Boyce, Madlen Witt, Martin Wilkinson, Brigit Hogan and Jimmy Kelly for helping provide many of the photographs included. I am grateful too to John Wiley & Sons for their professionalism, efficiency and never-ending patience as I spectacularly over-ran my manuscript submission date.

Gary Walsh
Limerick, June 2013

About the companion website

This book is accompanied by a companion website:

www.wiley.com/go/walsh/proteinsbiochemistry

The website includes:

- Powerpoints of all figures from the book for downloading
- PDFs of all tables from the book for downloading

Contents

此为试读,需要完整PDF请访问: www.ertongbook.com

Chapter 1
Proteins and proteomics

Throughout this book, I will consider various aspects of protein structure, function, engineering and application. Traditionally, protein science focused on isolating and studying one protein at a time. However, since the 1990s, advances in molecular biology, analytical technologies and computing has facilitated the study of many proteins simultaneously, which has led to an information explosion in this area. In this chapter such proteomic and related approaches are reviewed.

1.1 Proteins, an introduction

While we consider protein structure in detail in Chapter 2, for the purposes of this chapter it is necessary to provide a brief overview of the topic. Proteins are macromolecules consisting of one or more polypeptide chains (Table 1.1). Each polypeptide consists of a chain of amino acids linked together by peptide (amide) bonds. The exact amino acid sequence is determined by the gene coding for that specific polypeptide. When synthesized, a polypeptide chain folds up, assuming a specific three-dimensional shape (i.e. a specific conformation) that is unique to the protein. The conformation adopted depends on the polypeptide's amino acid sequence, and this conformation is largely stabilized by multiple, weak interactions. Overall, a protein's structure can described at up to four different levels.

- *Primary structure*: the specific amino acid sequence of its polypeptide chain(s), along with the exact positioning of any disulfide bonds present.
- *Secondary structure*: regular recurring arrangements of adjacent amino acid residues, often over relatively short contiguous sequences within the protein backbone. The common secondary structures are the α-helix and β-strands.
- *Tertiary structure*: the three-dimensional arrangement of all the atoms which contribute to the polypeptide. In other words, the overall three-dimensional structure (conformation) of a polypeptide chain, which usually contains several stretches of secondary structure interrupted by less ordered regions such as bends/loops.
- *Quaternary structure*: the overall spatial arrangement of polypeptide subunits within a protein composed of two or more polypeptides.

Proteins: Biochemistry and Biotechnology, Second Edition. Gary Walsh.
© 2014 John Wiley & Sons, Ltd. Published 2014 by John Wiley & Sons, Ltd.
Companion Website: www.wiley.com/go/walsh/proteinsbiochemistry

Table 1.1 Selected examples of proteins. The number of polypeptide chains and amino acid residues constituting the protein are listed, along with its molecular mass and biological function.

Protein	Polypeptide chains	Total no. of amino acids	Molecular mass (Da)	Biological function
Insulin (human)	2	51	5800	Complex, but includes regulation of blood glucose levels
Lysozyme (egg)	1	129	13,900	Enzyme capable of degrading peptidoglycan in bacterial cell walls
Interleukin-2 (human)	1	133	15,400	T-lymphocyte-derived polypeptide that regulates many aspects of immunity
Erythropoietin (human)	1	165	36,000	Hormone which stimulates red blood cell production
Chymotrypsin (bovine)	3	241	21,600	Digestive proteolytic enzyme
Subtilisin (*Bacillus amyloliquefaciens*)	1	274	27,500	Bacterial proteolytic enzyme
Tumour necrosis factor (human TNF-α)	3	471	52,000	Mediator of inflammation and immunity
Haemoglobin (human)	4	574	64,500	Gas transport
Hexokinase (yeast)	2	800	102,000	Enzyme capable of phosphorylating selected monosaccharides
Glutamate dehydrogenase (bovine)	~40	~8300	~1,000,000	Enzyme that interconverts glutamate and α-ketoglutarate and NH_4^+

The majority of proteins derived from eukaryotes undergo covalent modification either during, or more commonly after, their ribosomal synthesis. This gives rise to the concept of co-translational and post-translational modifications, although both modifications are often referred to simply as post-translational modifications (PTMs), and such modifications can influence protein structure and/or function. Proteins are also sometimes classified as 'simple' or 'conjugated'. Simple proteins consist exclusively of polypeptide chain(s) with no additional chemical components being present or being required for biological activity. Conjugated proteins, in addition to their polypeptide components, contain one or more non-polypeptide constituents known as prosthetic groups. The most common prosthetic groups found in association with proteins include carbohydrates (glycoproteins), phosphate groups (phosphoproteins), vitamin derivatives (e.g. flavoproteins) and metal ions (metalloproteins).

1.2 Genes, genomics and proteomics

The term 'genome' refers to the entire complement of hereditary information present in an organism or virus. In the overwhelming majority of cases it is encoded in DNA, although some viruses use RNA as their genetic material. The term 'genomics' refers to the systematic study of the entire genome of an organism. Its core aims are to:

- sequence the entire DNA complement of the cell; and
- to physically map the genome arrangement (assign exact positions in the genome to the various genes and non-coding regions).

Prior to the 1990s, the sequencing and study of a single gene represented a significant task. However, improvements in sequencing technologies and the development

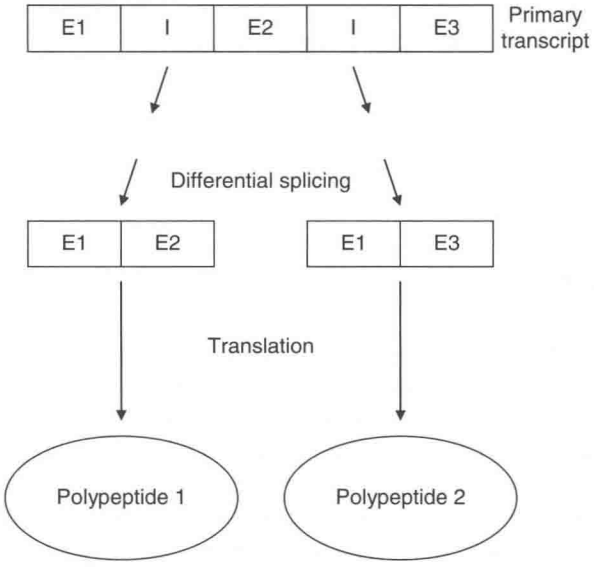


Figure 1.1 Differential splicing of mRNA can yield different polypeptide products. Transcription of a gene sequence yields a 'primary transcript' RNA. This contains coding regions (exons) and non-coding regions (introns). A major feature of the subsequent processing of the primary transcript is 'splicing', the process by which introns are removed, leaving the exons in a contiguous sequence. Although most eukaryotic primary transcripts produce only one mature mRNA (and hence code for a single polypeptide), some can be differentially spliced, yielding two or more mature mRNAs. The latter can therefore code for two or more polypeptides. E, exon; I, intron.

of more highly automated hardware systems now renders DNA sequencing considerably faster, cheaper and more accurate. Cutting-edge sequencing systems now in development are claimed capable of sequencing small genomes in minutes, and a full human genome sequence in a matter of hours and for a cost of approximately $1000. By early 2014, the genomes online database (GOLD; www.genomesonline.org), which monitors genome studies worldwide, documented some 36,000 ongoing/complete genome projects, and the rate of completion of such studies is growing exponentially. From the perspective of protein science, the most significant consequence of genome data is that it provides full sequence information pertinent to every protein the organism can produce.

The term 'proteome' refers to the entire complement of proteins expressed by a specific cell/organism. It is more complex than the corresponding genome in that:

- at any given time a proportion of genes are not being expressed;
- of those genes that are expressed, some are expressed at higher levels than others;
- the proteome is dynamic rather than static because the exact subset of proteins expressed (and the level at which they are expressed) in any cell changes with time in response to a myriad of environmental and genetic influences;
- for eukaryotes, a single gene can effectively encode more than one polypeptide if its mRNA undergoes differential splicing (Figure 1.1);
- many eukaryotic proteins undergo PTM.

The last two points in particular generally sigify that the number of proteins comprising a eukaryotic organism's proteome can far exceed the number of genes present in its genome. For example, the human genome comprises approximately 22,000 genes whereas the number of distinct protein structures present may exceed 1 million, with any one cell containing an estimated average of approximately 10,000 proteins.

Traditionally, proteins were identified and studied one at a time (Figure 1.2) (see Chapters 2, 3 and 4). This generally entailed purifying a single protein directly from a naturally producing cellular source,

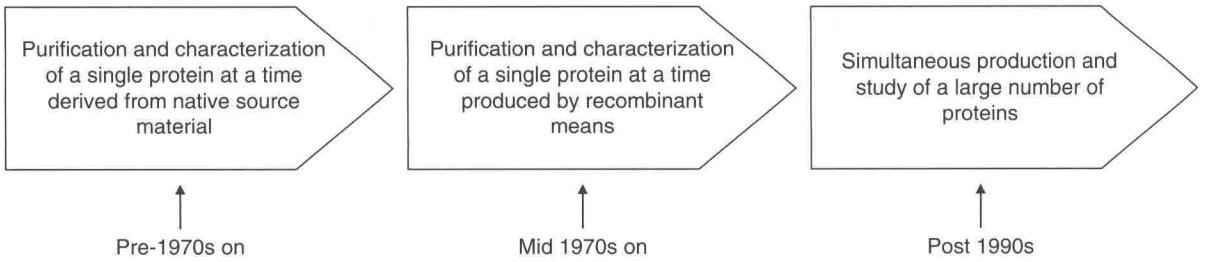


Figure 1.2 Evolution of the various approaches used to study proteins. Refer to text for details.

or from a recombinant source in which the gene/cDNA coding for the protein was being expressed. While this approach is still routinely used, a proteomic approach can potentially yield far more 'global' protein information far more quickly.

Proteomics refers to the large-scale systematic study of the proteome or, depending on the research question being asked, a defined subset of the proteome, such as all proteome proteins that are phosphorylated or all the proteome proteins that increase in concentration when a cell becomes cancerous. It is characterized by the integrated study of hundreds, more usually thousands or even tens of thousands of proteins. This in turn relies on high-throughput techniques/processes that facilitate the production, purification or characterization of multiple proteins rapidly and near simultaneously, usually by using automated/semi-automated and miniaturized processes/procedures. Standard techniques of molecular biology, for example, allow convenient global genome protein production (Figure 1.3) as well as facilitating the attachment of affinity tags to the proteins (as discussed later in this chapter and in Chapter 4), thereby enabling high-throughput purification efforts. Proteomics relies most of all on techniques that allow high-throughput analysis of the protein complement under investigation. Among the more central techniques in this regard are two-dimentional electrophoresis, high-pressure liquid chromatography (HPLC) and mass spectrometry (MS).

Before we consider the goals and applications of proteomics in more detail, it is worth reviewing these analytical techniques. In the context of proteomics, they are often applied in combination to characterize a target proteome, with electrophoretic and/or HPLC-based methods initially used to separate individual constituent proteome proteins from each other, followed by MS-based analysis. These techniques can also be used for the detailed analysis of individual proteins characteristic of classical protein science studies or, for example, as part of a quality control process for commercial protein preparations such as biopharmaceuticals. Such applications will be considered further in later chapters.

1.2.1 Electrophoresis

Electrophoresis is an analytical technique that separates analytes from each other on the basis of charge. The technique involves initial application of the analyte mixture to be fractionated onto a supporting medium (e.g. filter paper or a gel) with subsequent activation of an electrical field. Each charged substance then moves towards the cathode or the anode at a rate of migration that depends on the ratio of charge to mass (i.e. the charge density) of the analyte as well as on any interactions with the support medium. As described in Chapter 2, proteins are charged species, with their exact charge density being dependent on their amino acid sequence.

The most common electrophoretic method applied to proteins is one-dimensional polyacrylamide gel electrophoresis (PAGE) run in the presence of the negatively charged detergent sodium dodecyl sulfate (SDS-PAGE), and is most often used to analyse protein purity (see Chapter 4). In the case of PAGE, migration occurs through a polyacrylamide gel, the average pore size of which is largely dependent on the concentration of polyacrylamide present. A sieving effect therefore also occurs during PAGE so that the rate of protein migration is influenced by its size/shape as well as charge density.

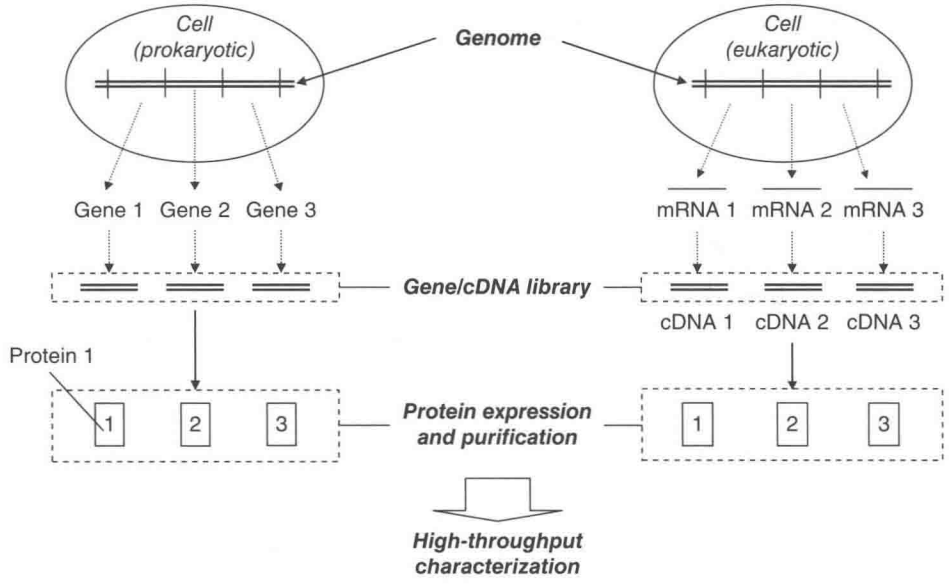


Figure 1.3 Global proteomics approach. While target proteins may be obtained from native (i.e. naturally producing) source material, they are most commonly obtained by recombinant means via the construction of gene/cDNA libraries. In the case of a prokaryotic cell source, a collection of individual genes can be isolated and cloned by standard molecular biology techniques, forming a genomic library (consisting of just three genes in the simplified example portrayed here). Eukaryotic genes generally consist of coding sequences (exons) interrupted by non-coding sequences (introns), while processed mRNA transcripts derived from those genes reflect the coding sequence for the final polypeptide product only. Isolation of total cellular mRNA followed by incubation with a reverse transcriptase enzyme yields complementary double-stranded DNA (cDNA) sequences, directly encoding the polypeptide sequences of the complement of expressed genes, thereby generating a cDNA library. Again by using standard molecular biology techniques the gene/cDNA library products can be expressed, yielding the recombinant protein products. The proteins, in turn, can be purified and characterized via techniques considered in subsequent sections of this chapter, as well as in Chapters 4 and 5.

Incubation of the protein with SDS has two notable effects: (i) it denatures most proteins, giving them all approximately the same shape, and (ii) it binds directly to the protein at the constant rate of approximately one SDS molecule per two amino acid residues. In practice this confers essentially the same (negative) charge density to all proteins. Separation of proteins by SDS-PAGE therefore occurs by a sieving effect, with the smaller proteins moving fastest towards the anode (Figure 1.4).

1.2.1.1 Isoelectric focusing

Isoelectric focusing is an additional form of electrophoresis. A modified gel is used which contains polyacrylamide to which a gradient of acidic and basic buffering groups are covalently attached. As a result an immobilized pH gradient is formed along the length of the gel. The gel is normally supported on a plastic strip. The protein solution to be applied is normally first incubated with a combination of urea and a non-ionic detergent such as Triton or CHAPS and a reducing agent to break any disulfide linkages present. This ensures that all sample proteins are completely disaggregated and fully solubilized. On application of the protein sample, the proteins present migrate in the gel until they reach a point at which the pH equals their isoelectric point (pI) (Figure 1.5).

Neither SDS-PAGE nor isoelectric focusing, by themselves, can fully separate (resolve) very complex mixtures of proteins, such as would characterize an entire cell's proteome. Each separation mode can individually resolve about 100 protein

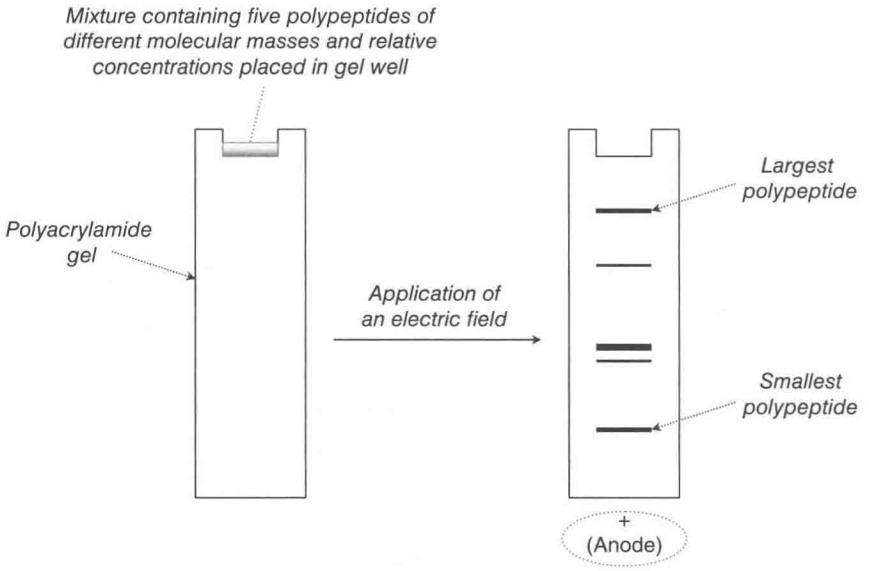


Figure 1.4 Separation of proteins by SDS-PAGE. Protein samples are incubated with SDS (as well as reducing agents, which disrupt disulfide linkages). The electric field is applied across the gel after the protein samples to be analysed are loaded into the gel wells. The rate of protein migration towards the anode depends on protein size. After electrophoresis is complete individual protein bands may be visualized by staining with a protein-binding dye.

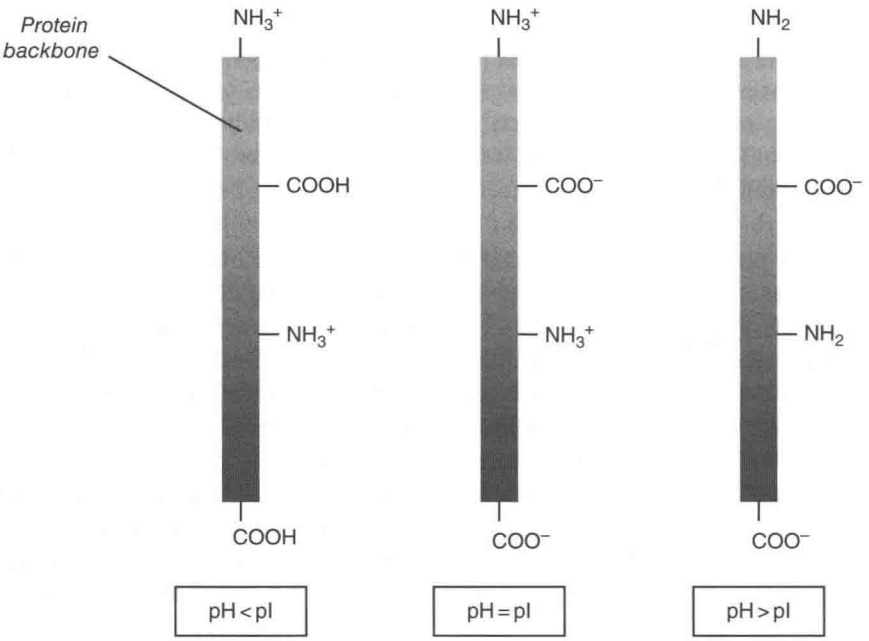


Figure 1.5 Proteins are amphoteric molecules, displaying a positive, negative or zero overall net charge depending on the pH of the solution in which are they dissolved. Contributing to the overall charge of a protein are all the positive and negative charges of its amino acid side chains as well as the free amino and carboxyl groups present at its amino and carboxyl termini, respectively. The state of ionization of these groups is pH dependent. The pH at which the net number of positive charges equal the net number of negative charges (i.e. the protein has an overall net electric charge of zero, and hence will not move under the influence of an electric field) is known as its isoelectric point (pI).

此为试读,需要完整PDF请访问: www.ertongbook.com