Molecular medical biochemistry

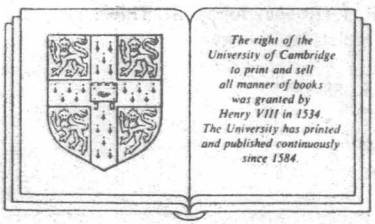
Molecular medical biochemistry

J. P. LUZIO

University Lecturer in Clinical Biochemistry,
University of Cambridge;
Fellow, St Edmund's College, Cambridge

R. J. THOMPSON

Professor of Clinical Biochemistry,
University of Southampton,
(sometime Fellow and Director of
Medical Studies, Corpus Christi College, Cambridge)



CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE
NEW YORK PORT CHESTER
MELBOURNE SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York NY 10011, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1990

Printed in Great Britain at the University Press, Cambridge

British Library cataloguing in publication data
Luzio, J. P.
Molecular medical biochemistry.
1. Medicine. Biochemistry
I. Title II. Thompson, R. J.
612'.015

Library of Congress cataloguing in publication data
Luzio, J. P.
Molecular medical biochemistry.
Includes index.
1. Clinical biochemistry. 2. Macromolecules – Metabolism. I. Thompson, R. J. (Rodney John) II. Title.
[DNLM: 1. Biochemistry. 2. Molecular Biology. 3. Proteins. QU 34 L979m]
RB112.5.L89 1990 612'.015 88-28[illegible]

ISBN 0 521 26083 3 hard covers
ISBN 0 521 27828 7 paperback

PREFACE

Biochemistry is taught to medical students at an early stage in their pre-clinical education. Often it seems less relevant to their future than other disciplines such as anatomy and physiology, and sometimes merely a course and examination hurdle to be overcome and quickly forgotten. Yet, despite the fact that it is biochemistry which so often loses teaching time to the encroachment of newer educational requirements (sociology, genetics, statistics, etc.), it is the equipment (centrifuges, scintillation counters, chromatography apparatus), and often ideas associated with biochemistry, not the accoutrements of anatomy or sociology, that fill the laboratories of departments of surgery and medicine. In accepting the opportunity to write this book, we felt we might persuade students that understanding recent developments in cellular and molecular biology will offer them as future clinicians the prospect of more rational diagnosis and therapy. In addition, we hope that a few students will be sufficiently enthused to consider research careers, applying the growing understanding of fundamental biochemical processes to a study of the common diseases of Man.

Writing even a short book requires a large amount of effort, help and advice from people other than the authors. We wholeheartedly thank our friends and colleagues, who have provided illustrations, read chapters, made useful comments and encouraged us in our endeavour. In addition, we thank Janet Eastwell, Lee Creswell and Diane Brown for typing the manuscript, and Fay Bendall and Alan Crowden of Cambridge University Press who never complained of the delays and inadequacies of our written material. We also thank those undergraduate students of the University of Cambridge whom we have supervised over the last decade (particularly at Pembroke, King's, Clare and Corpus Christi Colleges), and who have helped us inadvertently in the selection of topics for this book.

Finally, we thank especially Toni Luzio, who was extremely supportive throughout the writing.

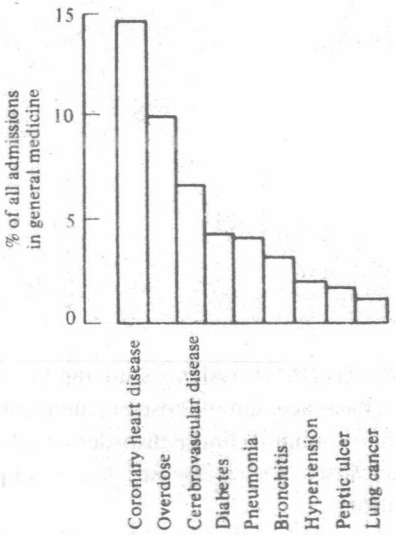
J. P. LUZIO
R. J. THOMPSON *Cambridge*

INTRODUCTION

This book is aimed at pre-clinical medical students studying biochemistry. In it, we have tried to give a basic account of those biochemical topics which we feel are most relevant to common clinical disorders in advanced industrial countries, even if the causes of such diseases do not yet permit complete biochemical explanation.

The health problems facing practising clinicians depend greatly on geographical location. World-wide there are 350 million cases of malaria per year (killing a million children in Africa alone), and at least 200–300 million people are infected with schistosomiasis, but neither of these diseases is prevalent in Western Europe. Moreover, the impact of infectious diseases has been greatly diminished in the past century by the development of good sanitation (and clean water supplies), vaccination programmes and the use of antibiotics. Nowadays, the common medical problems are exemplified by the histogram below which shows the main causes of general medical admissions to hospitals in England and Wales in a single year, with coronary artery disease, strokes, overdoses, diabetes and lung diseases of various sorts accounting for over half of the presenting clinical problems. It should be noted that the histogram excludes psychiatric admissions (a field where biochemistry, as opposed to pharmacology, has made little impact on diagnosis or therapy) and surgical admissions, where biochemistry is possibly least relevant.

The common diseases shown in the histogram are 'multifactorial', i.e. the end result of interaction between the products of many genes and environmental influences. The contribution of biochemistry to modern hospital practice in the treatment of such diseases is almost exclusively one of measurement, largely of substances in the serum or plasma of ill patients, sometimes on other body fluids or tissue samples and occasionally on well people. It is staggering to realize that an average-sized district hospital biochemical laboratory will carry out up to a million individual biochemical measurements per year, of which approximately 40% will



involve the use of an antibody or enzyme as a specific assay reagent. The majority of other measurements will be of simple ions or gases using physical methods. A significant proportion (10–20 %) of the work load will be quality control, i.e. the laboratory continually monitoring its own performance in the accuracy and precision of the measurements being made.

The contents of this book reflect the present use of analytical biochemical techniques to diagnose disease. Thus, in the early chapters, we describe some of the properties of proteins including antibodies and enzymes that make a particular contribution to the power of biochemical techniques in clinical diagnosis. Then we have selected topics (membrane structure and function, hormone action and secretion) where we believe current knowledge of macromolecular and cellular structure and function is likely to contribute most to understanding the underlying pathology of common multifactorial diseases. Finally, we have tried to outline the basic features of recombinant DNA techniques and the impact they are having in the diagnosis, treatment and knowledge of disease processes.

Throughout this book we assume that the reader will have access to one of the many comprehensive pre-clinical biochemistry textbooks. However, we hope we have written sufficiently clearly to preclude the need for constant reference to such a text. It is also our hope that the

reader becomes aware of the excitement of recent developments in cell and molecular biology and their potential impact on the practice of medicine.

CONTENTS

1

Proteins in medicine

Protein structure, including protein folding and domains

The human genome contains sufficient information to code for between 30 000 and 50 000 proteins. However, the majority of these (90–95 %) remain undiscovered and only a few per cent of the possible total number of human proteins are expressed at any one time in a particular cell type. Some of these proteins perform a structural role (e.g. those forming the cytoskeleton), others are enzymes catalysing and regulating metabolic pathways (e.g. the enzymes involved in glycolysis). The complement of proteins expressed by a particular cell determines its shape and functional capacity. How the diversity of cell types is directed and regulated (differentiation) remains a central problem in biology.

Proteins are large molecules varying in molecular weight from 1 to 1000 kDa and containing a linear sequence of amino acids covalently linked by peptide bonds. There is no clear dividing line between a peptide (a short chain of two or more amino acids) and a protein (a larger chain of amino acids usually with 100 amino acids or more). Chains of two or three amino acids are termed dipeptides or tripeptides, respectively, longer chains are often termed oligopeptides, and longer chains still polypeptides. Albumin in serum has 584 amino acids in a single linear chain; insulin has 51 amino acids contained in two polypeptide chains. All of these amino acids (with the exception of proline which is an imino acid) have the basic formula shown in Fig. 1.1(*a*) and differ only in the nature of the R group or side-chain. The 20 amino acids commonly found in proteins are listed in Fig. 1.2, together with the abbreviated symbols used to represent them. Amino acids occur naturally as *D*- or *L*- optical isomers, although only *L*-isomers are found in proteins. The nature of the R group determines the physical properties of the amino acid, and the polarity of the R group at pH 7.0 (which is close to physiological pH values) is particularly important. Some amino acids have negatively charged R groups at this pH (glutamic, aspartic), some have positively

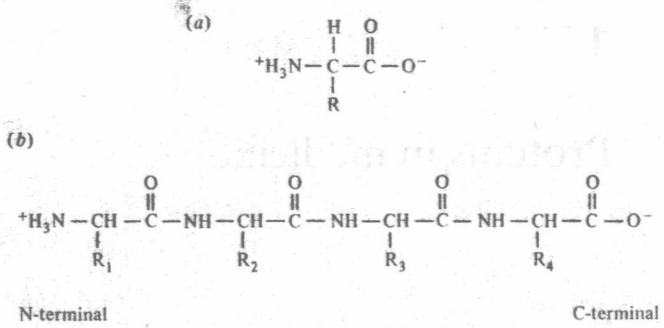


Fig. 1.1. (*a*) The basic structure of an amino acid. The R group varies according to the specific amino acid. At physiological pH the amino and carboxyl groups are charged as shown. (*b*) Amino acids are polymerized via peptide bonds between the amino group of one amino acid and the carboxyl group of the preceding amino acid. A tetrapeptide is shown. Note that polypeptide chains therefore have a free amino group at one end (the N-terminal end) and a free carboxyl group at the other end (the C-terminal end).

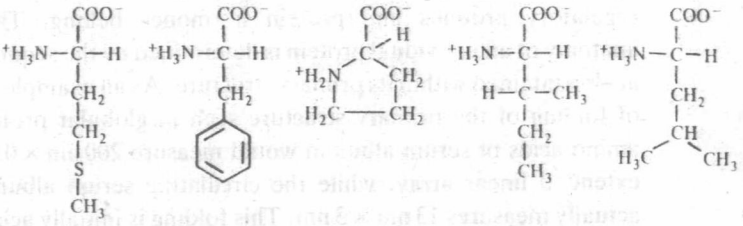
charged R groups (arginine, lysine, histidine), and others have uncharged R groups (glycine, see Fig. 1.2). The proportion of 'acidic' amino acids (negatively charged) (i.e. content of glutamate and aspartate) versus 'basic' amino acids (arginine, lysine, histidine) (positively charged) contained within a protein determines its overall charge at physiological pH, the total number of amino acids in the chain determines the overall molecular weight of the protein, and the order of amino acids determines the sequence or 'primary structure' of the protein. The 20 amino acids occurring in proteins can be arranged in a vast number of different sequences: for instance, a small polypeptide with each known amino acid occurring once can be arranged in 10^{18} different sequences; taking a protein of a relatively modest molecular weight of 35 kDa containing the 20 amino acids in equal numbers, and assembling one molecule of each possible sequence would produce a mass of protein greater than the weight of the earth.

Proteins exist in solution and within cells not as disordered linear sequences of amino acids but rather as folded structures with a molecular 'anatomy' characteristic for each protein. Two major 'anatomical' classes of protein are recognized, 'globular' proteins and 'fibrous' proteins. The latter, such as keratin found in hair and collagen found in tendons, are long, stringy, flexible molecules and are usually water-insoluble. Globular proteins, however, are tightly folded, water-soluble structures

and it is to this class that nearly all enzymes, antibodies, serum proteins, regulatory proteins and protein hormones belong. The molecular anatomy of an individual protein is determined by the sequence of amino acids contained within its primary structure. As an example of the degree of folding of the primary structure seen in globular proteins, the 584 amino acids of serum albumin would measure 200 nm × 0.5 nm if in an extended linear array, while the circulating serum albumin molecule actually measures 13 nm × 3 nm. This folding is initially achieved by two structural motifs, the α-helix and the β-pleated sheet (both of which were first discovered as structural features of fibrous proteins). The α-helix (Fig. 1.3) forms because of the partially double-bond character of the C–N bond which cannot rotate freely, while the adjacent C–C and N–C bonds can (Fig. 1.3). It also allows hydrogen bonding between the H atom attached to the electronegative N atom of the peptide linkage and the O atom of the carboxyl group of the fourth amino acid behind it (Fig. 1.3). The 'repeat unit' of the α-helix is a single turn of 0.54 nm, representing 3.6 amino acid residues. While the α-helix represents a stable preferred conformation for a chain of amino acids linked by peptide bonds, some constraints on its formation are provided by the particular amino acids involved. For example, proline cannot take part in α-helix formation since its N atom is part of a rigid ring and cannot rotate, and no hydrogen bond is possible between proline and a partner amino acid in the α-helix since the N atom in a proline ring does not have a substituent hydrogen atom. Proline therefore always interrupts an α-helix and produces a change of direction in the polypeptide chain. Highly charged amino acids (whether negatively charged, as with glutamic or aspartic, or positively charged, as with arginine or lysine) can also prevent α-helix formation if they are repeated too frequently in the amino acid sequence because of the electrostatic repulsion by like charges. Certain other amino acids, such as leucine and threonine, also tend to prevent α-helix formation if they occur too frequently in a short stretch of sequence because of the size and shape of their R groups.

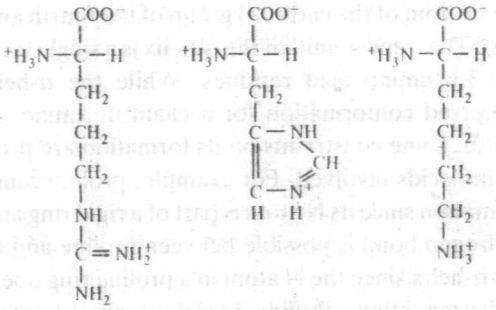
The second major arrangement of the primary structure of protein into an immediately higher-order structure is the β-pleated sheet. Here the polypeptide chain is more extended than in the α-helix and there are no intra-chain hydrogen bonds but rather inter-chain hydrogen bonds (Fig. 1.4). β-pleated sheets can form only in parts of the amino acid sequence where the R groups are relatively small. Proteins with extensive β-pleated sheet structures tend therefore to have high contents of glycine and alanine. The inter-chain hydrogen bonds between β-pleated sheets can be found between stretches of amino acid sequence running either in the

Non-polar (hydrophobic) R groups



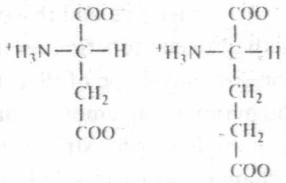
Methionine (Met or M) | Phenylalanine (Phe or F) | Proline (Pro or P) | Isoleucine (Ile or I) | Leucine (Leu or L)

Positively charged R groups



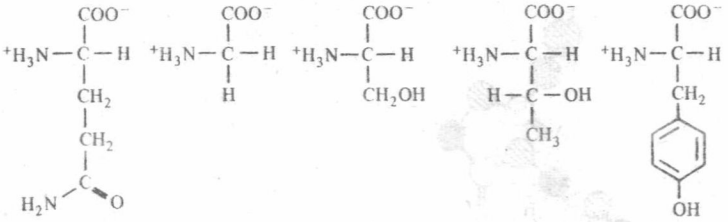
Arginine (Arg or R) | Histidine (His or H) | Lysine (Lys or K)

Negatively charged R groups



Aspartic acid (Asp or D) | Glutamic acid (Glu or E)

Polar but uncharged R groups



Glutamine (Gln or Q) | Glycine (Gly or G) | Serine (Ser or S) | Threonine (Thr or T) | Tyrosine (Tyr or Y)

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Fig. 1.2. The common amino acids found in proteins. The table shows the names of the amino acids with a three-letter code which is now being superseded by the one-letter code also shown. Examples of non-polar (hydrophobic), polar (uncharged), and polar (charged) groups are shown and also the structure of proline, which is an imino acid.

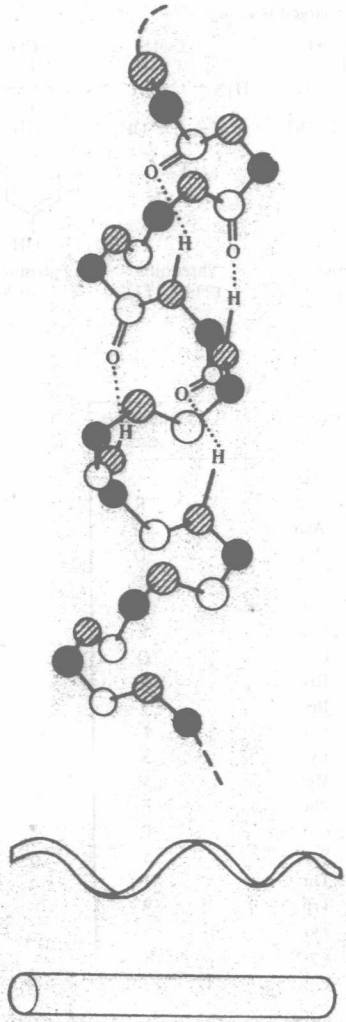


Fig. 1.3. The structure of the α-helix. Note the intra-chain hydrogen bond between the hydrogen atom attached to the nitrogen atom taking part in the peptide bond and the oxygen atom in the carboxyl group in the fourth amino acid above it. Here the R groups, which are on the outside of the α-helix, have been omitted. ● α-carbon atoms, ○-β-carbon atoms, ▨ nitrogen atoms. α-helices can be shown in diagrammatic representations of protein structures as spirals or cylinders.