The Biochemical Genetics of Man

edited by D. J. H. Brock
and O. Mayo

THE BIOCHEMICAL GENETICS OF MAN

edited by

D. J. H. BROCK
OLIVER MAYO

Department of Human Genetics, University of Edinburgh, Scotland

1972

Academic Press · London · New York

CONTRIBUTORS

D. R. Ackerman, *Department of Anthropology, University of California, Los Angeles, California, U.S.A.*

A. C. Allison, *Clinical Research Centre, Harrow, Middlesex, England.*

Gunhild Beckman, *Department of Clinical Bacteriology, University of Umeå, Umeå, Sweden.*

Barry Boettcher, *School of Biological Sciences, Flinders University of South Australia, Bedford Park, South Australia.*

D. J. H. Brock, *Department of Human Genetics, Western General Hospital, University of Edinburgh, Scotland.*

D. W. Cooper, *Department of Genetics, La Trobe University, Bundoora, Victoria, Australia.*

David M. Dawson, *Peter Bent Brigham Hospital, Boston, Massachusetts, U.S.A.*

K. W. E. Denson, *Research Laboratory, Oxford Haemophilia Centre, Churchill Hospital, Headington, Oxford, England.*

George Fraser, *Instituut voor Anthropogenetica, Rijksuniversiteit te Leiden, Leiden, Holland.*

Winfrid Krone, *Institut für Humangenetik und Anthropologie der Universität Freiburg, Freiburg, Germany.*

Oliver Mayo, *Department of Human Genetics, Western General Hospital, University of Edinburgh, Scotland.**

Wolfgang R. Mayr, *Institüt fur Blutgruppenserologie der Universität Wien, Vienna, Austria.*

M. W. Turner, *Department of Immunology, Institute of Child Health, University of London, England.*

J. M. White, *Department of Haematology, Royal Postgraduate Medical School, Hammersmith Hospital, London, England.*

Ulrich Wolf, *Institut für Humangenetik und Anthropologie der Universität Freiburg, Freiburg, Germany.*

---

* *Present address:* Biometry Section, Waite Agricultural Research Institute, University of Adelaide, Glen Osmond, South Australia.

PREFACE

A definition of biochemical genetics in a current textbook of human or medical genetics is difficult to find. This is surprising in view of the tremendous outpouring of work in the field and Professor Harry Harris's excellent summary of the principles of the subject ("The Principles of Human Biochemical Genetics", North Holland Publishing Company, 1970). It is also regrettable because biochemical genetics do represent a comparatively unified attempt to analyse human genetic variation at the molecular or near-molecular level. Because man is complex the types of variation which have been tackled are usually Mendelian ones; because DNA molecules are difficult to manipulate, the molecular level has meant the level of protein or enzyme (or their immediate products as in the blood groups and leucocyte antigens). Advances have been rapid and striking and it is no longer valid to dismiss biochemical genetics as those few aspects of human genetic variation which do not conveniently fit into more conventional categories.

In this book we have tried, with the help of our contributors, to be comprehensive in approach both to the fact and theory of human biochemical genetics. This is an ambitious aim, for some subjects (like blood group substances, leucocyte antigens and immunoglobulin variation) have become sciences in their own right. Coverage here must concentrate on recent advances and assume a certain basic knowledge (or willingness to acquire it) on the part of the reader. Nonetheless these topics have been included, for an account of biochemical genetics without them would seem to us to be quite misleading in emphasis.

The main body of the book is divided into two parts—common variants and rare variants. This has been convenient in ordering the chapters in some kind of logical sequence. But it is not a division based on principle, as the occurrence of many variants (e.g. those of the enzyme, glucose-6-phosphate dehydrogenase) in both parts will make clear. Indeed our aim in compiling this text has been to illustrate and emphasize through documentation of the facts the essential unity of biochemical genetics. We have been greatly helped by discussion with

numerous colleagues; in particular we thank Dr. Charles Smith and Dr. J. J. Veltkamp for their help with some of the manuscripts.

Edinburgh
June, 1972

D. J. H. Brock
Oliver Mayo*

---

* *Present address:* Biometry Section, Waite Agricultural Research Institute, University of Adelaide, Glen Osmond, South Australia.

CONTENTS

II. Normal Variation

I. Genetic Basis of Variation

1

THE STRUCTURE AND FUNCTION OF PROTEINS

D. M. DAWSON

Peter Bent Brigham Hospital, Boston, Massachusetts, U.S.A.

I. CLASSIFICATION OF PROTEINS

PROTEINS may be classified in a number of ways. Perhaps the most natural way to think of them is according to their function, and this is

done in Table I. Such a classification is useful, but scarcely does justice to the diversity of proteins, which matches the diversity of species. Another method derives historically from the solubility of proteins in aqueous solutions; there are soluble or globular proteins and insoluble or fibrous proteins. But not all globular proteins are freely soluble, especially storage proteins, and there are many lipoproteins and glycoproteins which are no more soluble than their prosthetic groups. In addition, the term globular protein has come to have a structural implication, namely a protein with optical properties (rotatory dispersion and circular dichroism) resembling those known for enzymes.

TABLE I[a]

FUNCTIONAL TYPES OF PROTEINS

Type	Function	Examples
1. Enzymes	Catalysis	Ribonuclease Carboxypeptidase Lactic dehydrogenase
2. Transport Proteins	Transport	Haemoglobin Serum albumin Serum lipoprotein
3. Storage Proteins	Storage	Ovalbumin Casein Zein
4. Antibody Proteins	React to foreign materials	Gamma globulin Complement
5. Structural Proteins	Support ? Transmembrane transport	Keratin Collagen Elastin Glycoproteins Fibrin(ogen) Permeases
6. Contractile Proteins	Movement	Myosin Actin Flagellar protein
7. Regulator Proteins	Regulation	Repressor protein (*E. coli*) ATCase subunit Insulin, ACTH Histones

[a] Modified from Lehninger (1970), p. 61.

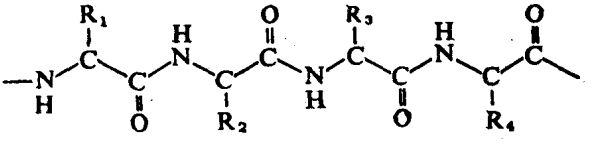
Such difficulties lead directly to a consideration of the various kinds of proteins and to a description of their known structural features. The object of this first chapter will be to present only examples of these main points and for this reason it will not claim in any sense to be comprehensive. Since much of the information on protein structure and function comes from organisms other than man, attention to the principles illustrated rather than exhaustive documentation of facts is considered relevant to the purposes of this book. A number of recent reviews deal more completely with certain of these areas and these will be referred to.

II. GENERAL FEATURES OF PROTEIN STRUCTURE

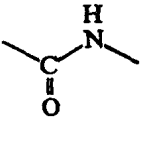
The primary structure of a protein is its amino acid composition or the moles of each amino acid per mole of protein. The secondary structure is a term largely used with reference to fibrous proteins, indicating the form or configuration of the peptide backbone. The tertiary structure usually refers to globular proteins and to their complete detailed three-dimensional structure. Quaternary structure means the arrangement of sub-units of multi-membered protein chains in complex molecules. The terms are artificial and are less useful now that it is realized that all are a direct result of the coded sequence of amino acids (see Lehninger, 1970).

A. The Polypeptide Chain

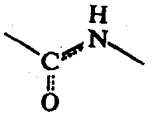
All proteins are linear polymers of peptide-linked amino acids, and have the general structure:



in which R_1 R_2 R_3 etc. refer to the side chains of the various amino acids. The representation given is one of the standard conventional illustrations, and does not really reveal the structure. The important feature is the



bond system (Pauling, 1960). The electrons present in the carbonyl group are in resonance across the C–N bond: statistically they "spend time" in the C–N bond as well as in the C–O bond, thus shortening the C to N distance. This resonance produces stabilisation of the unit and prevents rotation. (The property is a general one of amide groups.) The peptide structure could then be better represented as



It will be noted that the carbon groups on which R_1, R_2 etc. are substituted do not have partial double-bond character and rotation can occur there. These are the alpha-carbons and in a peptide bond they occur at a distance of 3·6 Å from each other.

The polypeptide chain thus consists of *repeating planar units*. Every 3·6 Å, at the R group-bearing carbon, rotation is allowed. Rotation is not or may not be random, but is determined by the size, charge and other properties of the R groups which are attached to the chain. Proteins in which the only organisation is that furnished by the polypeptide backbone are spoken of as *random coils*, although in many cases this usage does not indicate that the structure is random, but rather that we do not understand the functional significance conferred by that sequence of amino acids. The term random coil is also used in reference to proteins in which all known structure other than that of the polypeptide chain has been abolished, for instance, by dissolving them in 8M urea.

The peptide bond is formed in biological materials by a peptidyl synthetase present in the ribosomes and requires energy ultimately derived from adenosine triphosphate (ATP). It is a stable covalent bond and requires drastic chemical conditions to be hydrolysed. Commonly 6N HCl at a temperature of 110°C is used when full hydrolysis is desired. The peptide bond may also be broken by a number of enzymes (trypsin, chymotrypsin, subtilisin, papain etc.) and in a few cases the R group imparts special features which allow for hydrolysis of only the adjacent bond. Cyanogen bromide, for instance, cleaves the chain adjacent to methionyl residues. The standard free energy change for the hydrolysis of one peptide bond is −450 cal./mole, far less than that of a high energy phosphate compound, but fully enough to ensure that the rate of spontaneous hydrolysis of the peptide bond at neutral pH and 37°C will be vanishingly small.

B. The Alpha-Helix*

The earliest X-ray diffraction results on proteins more than 35 years ago disclosed that alpha keratin from wool fibres has a regular repeating structure with a periodicity somewhere between 5·0 and 5·5 Å along the axis. To explain this, the important structure of the alpha-helix was deduced by Pauling and Corey (1951), who began with model compounds, and devised the most stable hydrogen-bonded structures. The alpha-helix is a tightly coiled spiral, right-handed in sense, in which the amino-group hydrogen participates in a hydrogen bond with the carbonyl oxygen located four peptide bonds away. The R groups protrude out from the central axis. The rise along the axis of each turn of the helix gives a repeat period of 5·4 Å, agreeing well with that found for the wool fibres.

The helix itself achieves its stability from the hydrogen bonds formed parallel to the axis of the chain and can be built without any participation from the substituent R groups. This is not to say that the particular amino acids present in the peptide chain are without effect; some amino acids will block the formation of the helix, while others will not. Polylysine at neutral pH cannot form a helix, because of the adverse effect of the positively charged R groups; in an alkaline medium a polylysine helix will spontaneously form. Certain amino acids (e.g. serine) will tend to form other hydrogen bonds and do not favour helix formation. Proline, in which the N–C bond is part of a five-membered ring, cannot rotate at the alpha carbon and hence is unable at all to participate in alpha-helix formation, although polyproline makes two kinds of helices of its own.

C. Other Helices

The alpha-helix is not the only helical structure that can be formed by polypeptide chains. In the alpha-helix the hydrogen bonding is to the fourth carbonyl group up the chain, and the result is that there are 3·6 amino acid residues per turn of helix. Other helices include the π helix (bonding to the fifth carbon, 4·4 residues/turn) and the 3_{10} helix (bonding to the third carbon). To varying degrees the hydrogen bonds in these structures are subject to strain, the free energy is thereby reduced, and the helices are less stable. Proline, which cannot participate in the alpha-helix, can make its own helix as polyproline. This is

---

* The material in Sections B–D is largely derived from Dickerson and Geis (1969), which is a lively and well-illustrated introduction to protein structure, and it is also covered in Schellman and Schellman (1964).

not known to occur in natural proteins, nor is the left-handed alpha-helix which can be constructed from D-amino acids.

D. Beta-Conformation

It is also possible to stabilise proteins by forming hydrogen bonds between adjacent chains rather than along the axis of one chain. This is the basis of the β-conformation, which was also deduced by Pauling and Corey and called the *pleated sheet*. The peptide backbone in this case is a wavy or zigzag arrangement, and hydrogen bonds are formed between amino-group hydrogens and the carbonyl groups of parallel chains. When the peptide chains run in the same direction, they are said to have the parallel sense, and when in the opposite direction, the antiparallel sense. Both forms occur. The R groups project above and below each sheet and may interact with R groups from other sheets. Artificial polymers or copolymers of amino acids (for instance, polylysine) have again been important in studying these structures.

E. Covalent Bonds

Other than the peptide bonds, almost all covalent bonds found in proteins are those formed by the oxidation of two cysteine residues to form an –S–S-bridge. Their importance should not be over-estimated; many proteins do not contain any such cross-linkages. Some proteins contain no more than one or two such pairs of cysteines, and they appear to function more to orient the folding of chains. Serum proteins (albumin and γ-globulin) contain more than most.

F. Hydrophobic and van der Waal's forces

Particularly in the case of globular proteins, to be discussed below it, has become clear that much of the stability of the molecule is furnished by hydrophobic forces. These forces illustrate the vital influence of the R groups on protein structure. Some of the properties of the 20 amino acids coded for by DNA codons are listed in Table II. The hydrophobic amino acids in particular serve a special function when present in a protein chain. It will be appreciated that a polymer of leucine, for instance, will have only one positive charge at the amino-terminal and one negative charge at the carboxyl-terminal ends. Chains of neutral amino acids which are allowed to pack closely can exclude water from the environment of the individual amino acid side chains, producing hydrophobic regions. This process can be accompanied by the release