The principles of human biochemical genetics / Harry Harris.

FRONTIERS OF BIOLOGY

Edited by A. NEUBERGER AND E. L. TATUM

THE PRINCIPLES OF HUMAN BIOCHEMICAL GENETICS

HARRIS

THE PRINCIPLES OF HUMAN BIOCHEMICAL GENETICS

NORTH-HOLLAND RESEARCH MONOGRAPHS

FRONTIERS OF BIOLOGY

VOLUME 19

Under the General Editorship of

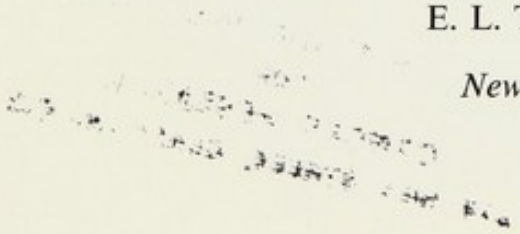
A. NEUBERGER

London

and

E. L. TATUM

New York



NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · LONDON

THE PRINCIPLES OF HUMAN BIOCHEMICAL GENETICS

HARRY HARRIS

Galton Laboratory, University College, London

1970

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · LONDON

Library of Congress Catalog Card Number: 75–108280
ISBN: 7204 7119 2

PUBLISHERS:

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM
NORTH-HOLLAND PUBLISHING COMPANY, LTD. – LONDON

SOLE DISTRIBUTORS FOR THE U.S.A. AND CANADA:

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

This book is also published in a paperback student edition

PRINTED IN THE NETHERLANDS

Previous volumes in the series **FRONTIERS OF BIOLOGY**

Volume 1: **Microbial models of cancer cells**
G. F. GAUSE, Academy of Medical Sciences, Moscow

Volume 2: **Interferons**
Edited by N. B. FINTER, Imperial Chemical Industries, Macclesfield, Cheshire

Volume 3: **The biochemical genetics of vertebrates except man**
I. E. LUSH, Rowett Research Institute, Aberdeen

Volume 4: **Delayed hypersensitivity**
J. L. TURK, University of London

Volume 5: **Human population cytogenetics**
W. M. COURT BROWN, Western General Hospital, Edinburgh

Volume 6: **Thymidine metabolism and cell kinetics**
J. E. CLEAVER, University of California

Volume 7: **The cell periphery, metastasis and other contact phenomena**
L. WEISS, Roswell Park Memorial Institute, Buffalo, N.Y.

Volume 8: **The electrostatics of biological cell membranes**
R. M. FRIEDENBERG, University of Maryland

Volume 9: **The pyrrolizidine alkaloids. Their chemistry, pathogenicity and other biological properties**
L. B. BULL, C. C. J. CULVENOR and A. T. DICK, Commonwealth Scientific & Industrial Research Organization, Australia

Volume 10: **Antagonists and nucleic acids**
M. EARL BALIS, Sloan Kettering Institute for Cancer Research, Rye, N.Y.

Volume 11: **Macrophages and immunity**
D. S. NELSON, Department of Bacteriology, The University of Sydney, Australia

Volume 12: **The biological code**
M. YČAS, Department of Microbiology, State University of New York, N.Y.

Volume 13: **The biochemistry of folic acid and related pteridines**
R. L. BLAKLEY, Department of Biochemistry, John Curtin School of Medical Research, Australian National University

Volume 14: **Lysosomes in biology and pathology** (complete in 2 volumes)
Edited by J. T. DINGLE and HONOR B. FELL, Strangeways Research Laboratory, Cambridge

Volume 15: **Handbook of molecular cytology**
Edited by A. LIMA-DE-FARIA, Institute of Molecular Cytogenetics, University of Lund, Sweden

Volume 16: **Immunogenetics of tissue transplantation**
ALENA LENGEROVÁ, Institute of Experimental Biology and Genetics, Czechoslovak Academy of Sciences, Prague

Volume 17: **Prenatal respiration**
HEINZ BARTELS, Physiologisches Institut der Medizinischen Hochschule, Hannover, Germany

Volume 18: **The transmission of passive immunity from mother to young**
F. W. ROGERS BRAMBELL, Emeritus Professor of Zoology, Bangor, Caerns., United Kingdom

Editors' preface

The aim of the publication of this series of monographs, known under the collective title of '*Frontiers of Biology*', is to present coherent and up-to-date views of the fundamental concepts which dominate modern biology.

Biology in its widest sense has made very great advances during the past decade, and the rate of progress has been steadily accelerating. Undoubtedly important factors in this acceleration have been the effective use by biologists of new techniques, including electron microscopy, isotopic labels, and a great variety of physical and chemical techniques, especially those with varying degrees of automation. In addition, scientists with partly physical or chemical backgrounds have become interested in the great variety of problems presented by living organisms. Most significant, however, increasing interest in and understanding of the biology of the cell, especially in regard to the molecular events involved in genetic phenomena and in metabolism and its control, have led to the recognition of patterns common to all forms of life from bacteria to man. These factors and unifying concepts have led to a situation in which the sharp boundaries between the various classical biological disciplines are rapidly disappearing.

Thus, while scientists are becoming increasingly specialized in their techniques, to an increasing extent they need an intellectual and conceptual approach on a wide and non-specialized basis. It is with these considerations and needs in mind that this series of monographs, '*Frontiers of Biology*' has been conceived.

The advances in various areas of biology, including microbiology, biochemistry, genetics, cytology, and cell structure and function in general will be presented by authors who have themselves contributed significantly to these developments. They will have, in this series, the opportunity of bringing together, from diverse sources, theories and experimental data, and of integrating these into a more general conceptual framework. It is unavoidable, and probably even desirable, that the special bias of the individual authors will become evident in their contributions. Scope will also be given for presentation of new and challenging ideas and hypotheses for which

complete evidence is at present lacking. However, the main emphasis will be on fairly complete and objective presentation of the more important and more rapidly advancing aspects of biology. The level will be advanced, directed primarily to the needs of the graduate student and research worker.

Most monographs in this series will be in the range of 200–300 pages, but on occasion a collective work of major importance may be included exceeding this figure. The intent of the publishers is to bring out these books promptly and in fairly quick succession.

It is on the basis of all these various considerations that we welcome the opportunity of supporting the publication of the series '*Frontiers of Biology*' by North-Holland Publishing Company.

E. L. TATUM
A. NEUBERGER, *Editors*

Preface

This book has grown out of a course of lectures given at the Galton Laboratory, which was intended not only for students and research workers in human genetics, but also for biochemistry, biology and medical students as well as for interested research workers in related fields. I was concerned to explain the principal concepts which underlie modern ideas in human biochemical genetics, to present a picture of the extraordinary degree of inherited biochemical diversity which recent research has shown to be a characteristic feature of human populations, and to show how the detailed analysis of genetically determined biochemical differences between individual members of our species could throw new light on fundamental problems not only in genetics, but also in medicine and more generally in human biology.

Just over ten years ago I wrote an account of the subject (Human Biochemical Genetics, Cambridge University Press, 1959) covering most of the information which was then available, in what seemed at the time a logical order. Since then, however, research in this field has expanded almost explosively and in preparing the present book, it became very obvious that one could not be content with simply trying to update the earlier text. The many advances now called for a very different arrangement if present knowledge and concepts were to be presented in a coherent manner. This is not merely because a great deal more is now known about the particular topics that were dealt with in the earlier book, but because whole new areas of the subject have been opened up in a manner which could hardly have been envisaged only a few years ago. One of the important consequences of these developments has been the greater unity they have given to the subject as a whole. The interrelationships between what at one time seemed very different and unconnected types of phenomena such as the inborn errors of metabolism, the blood group antigenic differences, the haemoglobin diseases and the enzyme and protein polymorphisms, can now be thought about within a consistent theoretical framework in a way that was hardly possible previously. This of course gives one an opportunity to try and present the subject in a more systematic and analytical manner. So the present work differs consider-

ably in its approach and arrangement from my earlier one, and of course much of the material discussed is new.

One of the difficulties in writing this kind of text is deciding what examples should be used to illustrate the various points in the argument, and in how much detail they should be presented. Also one has to decide what must be left out, if the overall length is to be kept to a manageable size, and the argument not be obscured by an excessive amount of descriptive material. Since a book of this sort may also be useful as a source of reference, one is often placed in something of a quandary. I have tried to resolve this difficulty by giving key references to much material which is not described in detail in the text, and a great deal of this has been arranged in the form of tables or appendices so that the appropriate references can be extracted more easily. Nevertheless the scope of the subject is now so very extensive and the literature so vast and distributed over such a wide range of journals, that reference to many topics must inevitably have been omitted. I hope nevertheless that enough has been included so that the book may serve not only as an introductory text in which the main principles of the subject are formulated, but also as guide to further reading on specific topics.

Galton Laboratory,
University College, London

HARRY HARRIS, M.D., F.R.S.
July, 1969

Contents

Acknowledgements

It is a great pleasure to thank Mrs. N. Parry-Jones for preparing the figures, Mrs. J. Barrie for her secretarial assistance, Dr. Moyra Smith for the index and Dr. D. A. Hopkinson for much helpful discussion and advice while the book was being written.

Many of the illustrations come from other publications and the permission for their reproduction is gratefully acknowledged. The original sources are given in the captions and biobliography.

CHAPTER 1

Gene mutations and single aminoacid substitutions

1.1. Introduction: genes, DNA and proteins

Human beings are exceedingly diverse. They differ from one another in their normal physical, physiological and mental attributes. They also differ in whether they suffer from particular diseases or other abnormalities. These variations are caused in part by differences in the environmental conditions in which they live. But they also depend on inborn differences. Indeed it is very probable that no two individuals with the exception of monozygotic twins are exactly alike in their inherited constitutions. The analysis in molecular terms of the nature and effects of such genetically determined differences forms the subject matter of human biochemical genetics.

Classical genetics led to the concept of the gene as the fundamental biological unit of heredity and postulated that it must possess three basic properties. It had to have a specific function in the cell, and hence in the organism as a whole. It had to be capable of exact self-replication so that its functional specificity would be preserved from one cell generation to the next. Finally, although usually an extremely stable entity, it had to be susceptible to occasional sudden change or mutation, which could result in the appearance of a new unit or allele differing functionally from the original one but self-replicating in its new form. It was shown how such units are arranged in linear order in the chromosomes, each gene having its own characteristic position or locus; how they are transmitted to an individual from his parents via the ovum and sperm, so that they are usually present in pairs, one member of a pair being derived from one parent and one from the other; and how because of mutational changes in previous generations multiple allelic forms of a gene can occupy a particular gene locus, so that individual members of a natural population may differ from one another in their characteristics according to the specific nature of the alleles that they happened to have received from their parents.

Four major advances made it possible to begin to understand the nature of

genetical diversity in molecular terms. The first was the discovery that the particular chemical substance which endows a gene with its characteristic properties is deoxyribosenucleic acid or DNA. The second was the elucidation of the molecular structure of this substance. The third was the recognition that the primary biochemical role of DNA in the cells of an organism is to direct the synthesis of enzymes and other proteins. The fourth was the unravelling of the genetic code. That is the details of the relationships between the structure of nucleic acids and the structure of proteins.

The main features of the molecular architecture of DNA were first formulated by Watson and Crick in 1953, who at the same time pointed out how the proposed structure would account for the three basic attributes of genetic material; gene specificity, gene replication and gene mutation.

The molecule is made up of two very long polynucleotide chains coiled round a common axis to form a double helix. The backbone of each chain consists of a regular alternation of phosphate and sugar (deoxyribose) groups. To each sugar group and projecting inwards from the chain is attached a nitrogenous base. This may be one of four different types; adenine or guanine which are purines, thymine or cytosine which are pyrimidines. The two chains are held together by hydrogen bonding between pairs of bases projecting at the same level from each chain, so that the whole structure may be likened to a spiral staircase, the pairs of bases representing the steps. There are certain restrictions on which bases can constitute a pair. In any one pair one base must be a purine and the other a pyrimidine, and of the possible combinations only two can occur; adenine with thymine, and guanine with cytosine. A given pair may be either way round. Thymine for example can occur in either chain, but when it does its partner on the other chain must be adenine.

A gene can be regarded as being represented by a length of DNA containing several hundred or thousand base pairs. While the phosphate–sugar backbones of the two chains which form the double helix are quite regular, the base pairs may occur in any sequence. A great many different permutations are therefore possible, and so each gene can have its own unique structure, from which is derived its functional specificity. The precise sequence of base pairs in a particular gene carries as it were in coded form a specific piece of genetical information.

Since the nature of one base fixes the nature of the other member of the pair, the two polynucleotide chains which make up the molecule, though qualitatively different, are exactly complementary. The sequence of bases in one chain fixes the sequence of bases in the other. Replication can occur by

the unwinding and separation of the chains and the reformation on each chain of its appropriate companion from an available pool of nucleotides. Each chain may thus act as a template for the formation of the other, so that from one molecule two precise replicates are produced each with exactly the same sequence of base pairs as the original.

A gene mutation can be envisaged as the consequence of some kind of event which results in an alteration of the base pair sequence of the particular gene. Many and perhaps most mutations probably represent no more than the change of one base for another at some point in the sequence. Others however involve more drastic changes such as the duplication or deletion of part of the sequence or some other kind of rearrangement. In general the new gene structure once formed will then be conserved in subsequent cell divisions by the ordinary process of DNA replication.

A great variety of different enzymes and other proteins are synthesised in the cells of a single organism. They each have their own distinctive properties and functions and together they define and control the complex pattern of metabolic and developmental processes which characterise the species and the individual. Proteins are composed of one or more polypeptide chains which are made up of long strings of aminoacids linked by peptide bonds in a specific linear order. Twenty different aminoacids may be present and typical polypeptide chains have sequences 100–500 aminoacids long, so as with DNA the number of possible structures is enormous. Furthermore the three-dimensional arrangement and hence the characteristic properties and functional activity of any given protein ultimately depends on the precise sequence of aminoacids in its constituent polypeptide chains.

The fundamental idea relating DNA structure to protein structure is that the sequence of base pairs in a given gene determines the sequence of amino-acids in a corresponding polypeptide chain. So the structures and hence the properties of all the enzymes and proteins an individual can make are thought to be defined by the base pair sequences of his genes.

The details of the genetic code – that is the relationship of base sequence to aminoacid sequence – have largely been worked out by experimental studies on microorganisms. But there is little doubt that in their main features they also apply in higher organisms including man. Each aminoacid is specified by a sequence of three bases. The base triplets occur consecutively and do not overlap. That is to say a triplet specifying one aminoacid is immediately followed by a separate triplet specifying the next aminoacid and so on. Thus the two sequences are colinear. The four characteristic bases of a DNA chain can occur in 64 different triplet sequences, and 61 of these

triplets each specify one of the twenty aminoacids, so that a particular aminoacid may be coded by two or more different base triplets (see fig. 1.5, p. 12). There are also three so-called 'nonsense' triplets which do not specify aminoacids but probably designate chain termination.

The series of processes by which the sequence of bases in the DNA of a gene is translated into the corresponding sequence of aminoacids in a polypeptide chain are complex and involve as intermediaries certain types of ribosenucleic acid (RNA) molecules. The first step involves the separation of the two polynucleotide chains of the DNA, so that one of them may serve as a template for the synthesis from available ribonucleotides of a complementary RNA chain. In this process the same base pairing rules as in DNA apply, except that uracil, which occurs in RNA instead of thymine, pairs with adenine. Thus a strand of RNA carrying the same genetic information as the DNA strand is formed, but it is coded in a complementary base sequence. This RNA strand, known as messenger or mRNA, then separates from the DNA and passes out of the cell nucleus to the ribosomes in the cytoplasm, which are the site of protein synthesis. The strands of mRNA attached to ribosomes then serve as templates for the formation of the polypeptide chains. Aminoacids come to the mRNA template attached to another species of RNA molecule, known as transfer RNA or tRNA. The tRNA molecules are relatively small (about 80 nucleotides) and occur as a series of distinct molecular types. Each is specific for a particular aminoacid which can be attached to one end of the tRNA molecule, and each contains within its polynucleotide sequence a characteristic base triplet complementary to a base triplet in mRNA which codes for that aminoacid. The attached aminoacid can thus be placed in the correct position for the synthesis of the polypeptide chain defined by the coded sequence of the mRNA.

The polypeptide chain is made sequentially, one aminoacid being added at a time, starting from the amino-terminal end. A key feature of the processes leading to polypeptide synthesis is that each aminoacid in the sequence is designated by a trinucleotide or base triplet in three types of molecule: DNA, mRNA and tRNA. The triplet in mRNA is complementary to that in DNA and also to that in tRNA, so that although the actual bases differ the same aminoacid is specified.

Thus the inherited information coded in the genes can be regarded as a kind of blueprint which defines the structures of all the enzymes and other proteins which an individual makes. But genes not only determine the structures of proteins, they are also apparently concerned in regulating their synthesis. The nature of the molecular relationships involved in these regula-

tory functions are still however very obscure, and so far no satisfactory general theory, at least for multicellular organisms, has as yet been developed (see ch. 4, p. 90). Nevertheless, one can say that the enzyme and protein makeup of each individual must in a very direct sense be a reflexion of his genetical constitution. Furthermore, one may anticipate that inherited differences between individual members of the species, whether they are expressed as differences in normal physical, physiological or mental characteristics, or as differences in the development of particular abnormalities, are likely to be a consequence of differences in enzyme or protein synthesis.

1.2. The haemoglobin variants

The first direct evidence that a gene mutation can result in the synthesis of an altered protein came from work on haemoglobin in the condition called sickle-cell disease or sickle-cell anaemia.

It has been known for many years that the red blood cells of certain individuals have the peculiar property of undergoing reversible alterations in shape when subjected to changes in the partial pressure of oxygen (Herrick 1910, Hahn and Gillespie 1927). When oxygenated, the cells are biconcave discs like the red cells of normal individuals. However, when they are deoxygenated, they become elongated, filamentous and sickle-shaped. Most people whose red cells show the so-called 'sickling phenomenon' are quite healthy. They are said to have the sickle-cell trait. In some individuals, however, 'sickling' is associated with a severe and characteristic form of anaemia, which is commonly fatal in childhood or adolescence. This condition is usually called sickle-cell disease. The sickling phenomenon is common in Central Africa where in many areas 20% or more of the population may have the sickle-cell trait, and a significant fraction of individuals (1–2%) may be expected to die of sickle-cell anaemia in early life. It is also found not infrequently among Negro populations living in other parts of the world, such as the U.S.A. But it is rare or absent in most other human populations.

The peculiarity is an inherited characteristic, and in 1949 Neel and Beet working independently showed that the pedigrees could be most simply explained by the hypothesis that individuals with the sickle-cell trait and those with sickle-cell anaemia are respectively heterozygous and homozygous for a particular abnormal gene located on one of the autosomal chromosomes. Thus individuals who receive the abnormal gene from one parent but its normal allele from the other would be expected to have the sickle-cell trait,

while those who receive the abnormal gene from each parent would develop sickle-cell anaemia. On this hypothesis one expects that in matings between a normal individual and one with the sickle-cell trait, half the children should on the average be normal and half should have the sickle-cell trait. From matings between two individuals with the sickle-cell trait, there should be normal children, children with the sickle-cell trait, and children with sickle-cell anaemia, and these should occur on average in the Mendelian ratio of 1:2:1. Furthermore both parents of a patient with sickle-cell anaemia should exhibit the sickling phenomenon. The family data (Neel 1951) were shown to be in good agreement with these and other expectations, and subsequent studies fully confirmed the hypothesis.

At about the same time Pauling and his colleagues (1949) made the crucial discovery that the haemoglobin present in the red cells in sickle-cell anaemia is qualitatively different from the haemoglobin present in normal red cells. They showed that the two proteins differed in their physical properties and presumably therefore in their structures, since they could be separated by electrophoresis. Furthermore they demonstrated that the red cells from individuals with the sickle-cell trait contained both normal haemoglobin (Hb A) and also the abnormal sickle-cell haemoglobin (Hb S). Thus there appeared to be a direct correspondence between the genetic constitution of an individual and the haemoglobins that were synthesised. Individuals homozygous for the sickle-cell gene formed Hb S, individuals homozygous for the normal allele formed Hb A, and heterozygous individuals who had inherited the sickle-cell gene from one parent and its normal allele from the other formed both types of haemoglobin.

Shortly afterwards another abnormal haemoglobin, called Hb C, was discovered (Itano and Neel 1950). Like Hb S this also appeared from family studies to be determined by a single gene. Heterozygotes for this gene and its normal allele form both Hb A and Hb C in their red cells. They are quite healthy and are said to have the haemoglobin C trait. Homozygotes for the gene form Hb C but no Hb A and may show a moderate or mild degree of anaemia. This condition, known as haemoglobin C disease, is clinically much less severe than sickle-cell disease. A third condition of intermediate severity also occurs. This is known as sickle-cell–haemoglobin C disease, and affected individuals have both Hb S and Hb C in their red cells. Family studies indicate that they are heterozygous for both the gene determining Hb C and also for the sickle-cell gene, having inherited one of these abnormal genes from one parent and one from the other. These various conditions can be readily

distinguished from each other by electrophoresis. Typical separations are illustrated in fig. 1.1.

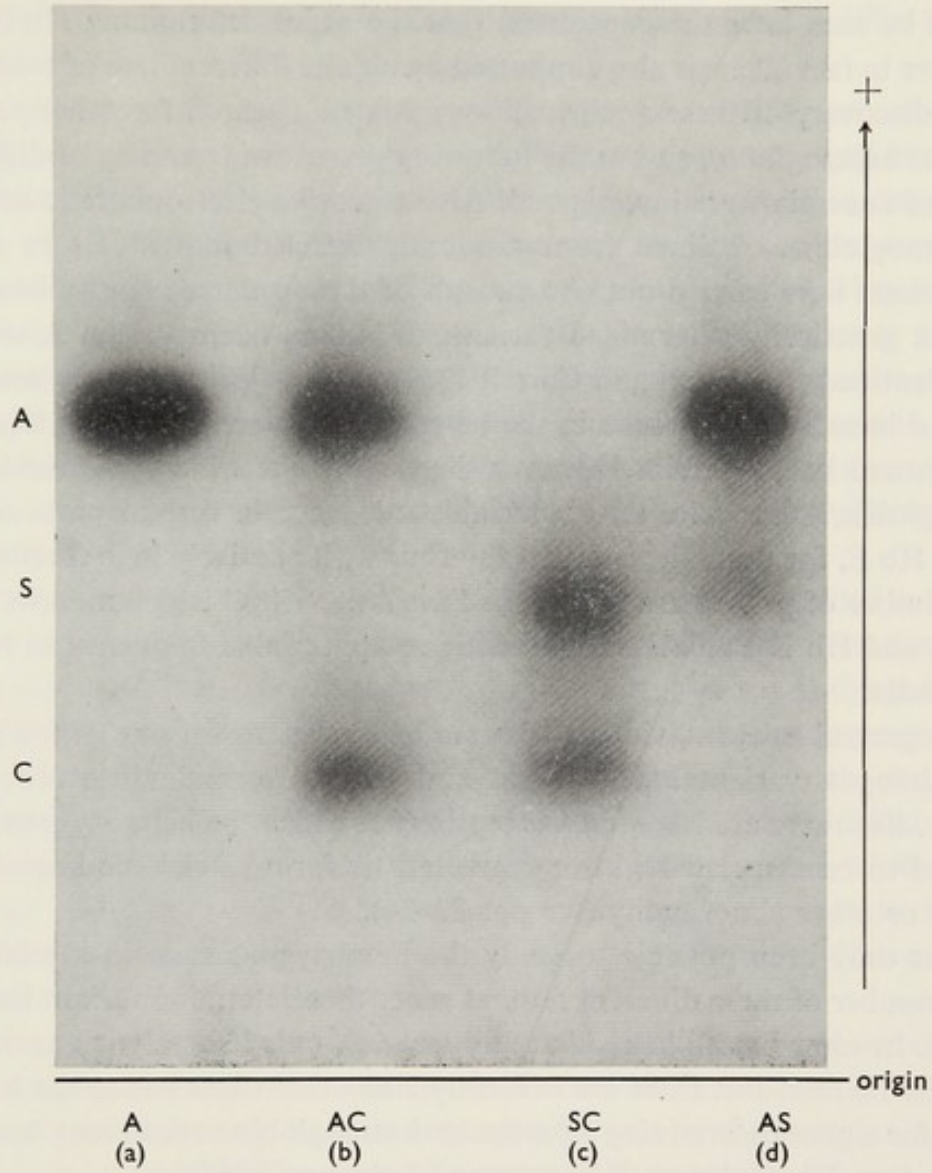


Fig. 1.1. Electrophoresis of haemoglobins in (a) normal adult, (b) haemoglobin C trait, (c) sickle-cell–haemoglobin C disease, and (d) sickle-cell trait. (Electrophoresis carried out at *p*H 8.6 in starch gel.)

In sickle-cell–haemoglobin C disease no Hb A is found. This suggests that the genes determining Hb S and Hb C are allelic, and that when both occur the normal allele presumed to be necessary for the formation of Hb A cannot be present. Studies of families in which one parent had sickle-cell–haemoglobin C disease while the other was normal proved to be consistent with this

idea. The offspring were found to include individuals with either the haemoglobin C trait (AC) or the sickle-cell trait (AS), but not individuals with sickle-cell–haemoglobin C disease (SC) or with normal haemoglobin (A). As will be seen later the conclusion that the genes determining Hb S and Hb C are in fact allelic is also supported by a quite different line of evidence.

The discovery of these abnormalities initiated a search for other variant forms of haemoglobin, and in the following years a wide variety of different forms of anaemia were investigated. Also extensive electrophoretic surveys of haemoglobins obtained from randomly selected individuals in many populations were carried out. As a result of these studies about a hundred different genetically determined variants of human haemoglobin have now been identified (Lehmann and Carrell 1969). The majority of these are very rare and have only been seen in the heterozygous state, occurring together with normal haemoglobin. However there are some which, like sickle-cell haemoglobin, have quite an appreciable incidence in certain parts of the world. Hb E, for example, is found to occur with relatively high frequencies in a number of populations in South East Asia, Hb C is common in West Africa, and Hb D Punjab is found with an appreciable frequency in North West India.

As a general rule, individuals who are heterozygous for one of the genes determining a variant haemoglobin and for its normal allele are quite healthy. But there are important exceptions in which the heterozygous state is found to be associated with a characteristic form of chronic haemolytic disease or other abnormality (see pp. 23–24).

It has only been possible to study the homozygous state in a relatively small number of these different mutant genes that determine variant haemoglobins. In some cases like sickle-cell disease chronic haemolytic anaemia is a regular feature. But there are evidently also other cases where the homozygote for a gene determining a particular haemoglobin variant may be quite healthy (e.g. Hb G Accra, Edington and Lehmann 1954).

Chronic haemolytic disease can also be a characteristic feature of certain heterozygous states like sickle-cell–haemoglobin C disease where two different abnormal alleles are present. It is also often seen in individuals who are heterozygous for one of the genes determining a particular variant haemoglobin, and also heterozygous for one of the genes to be discussed later (ch. 4) which result in a specific defect in haemoglobin synthesis (the so-called thalassaemia genes). Thus a variety of different haematological abnormalities which had previously not been clearly differentiated from one another can now be defined in terms of specific abnormalities in haemoglobin formation.

1.3. The structure of the variant haemoglobins

With the discovery of the many genetically determined variants of haemoglobin came the problem as to the precise way these proteins differ in molecular structure from normal haemoglobin.

A protein consists of one or more polypeptide chains each of which is composed of aminoacids linked by peptide bonds and arranged in a definite sequence. Where there are two or more polypeptide chains in the structure of a particular protein molecule, these may be identical in sequence or non-identical. Haemoglobin A has four polypeptide chains. They are of two different types, each with a characteristic aminoacid sequence and each represented twice in the molecule. These different polypeptides are called the α- and the β-chains, so the haemoglobin A molecule may be said to have the structure $\alpha_2\beta_2$. The α-chain contains 141 amino acid residues and the β-chain 146 amino acid residues, and their precise sequences (Braunitzer et al. 1964) have been established (see fig. 2.7, p. 36).

Each of the polypeptide chains in a protein is coiled and folded in a characteristic manner so that the whole molecule has a complex three-dimensional arrangement. The spatial configuration taken up by each of the polypeptide chains and the consequent overall three-dimensional arrangement of the protein molecule is thought to be determined primarily by the sequence of aminoacids in the constituent polypeptide chains, the so-called primary structure. In the case of haemoglobin, the detailed manner in which the four chains are coiled and folded and the way they fit together to form the globular protein molecule has been worked out by Perutz and his colleagues (1968) by X-ray crystallographic analysis.

Many proteins have attached to them some additional and usually relatively small groupings not made up of aminoacids. These are called prosthetic groups and they may have a very specific role in the functioning of the molecule as a whole. The prosthetic group in haemoglobin is haem. This is a porphyrin ring structure with an iron atom at its centre. There are four haem groups in the haemoglobin molecule, one being attached to each of the polypeptide chains by a linkage between the iron atom and a particular histidine residue in each of the chains. In the three-dimensional model of haeomoglobin the haem groups are seen to lie in four separate pockets formed by folds in the corresponding polypeptide chains. When oxygen combines with haemoglobin it attaches to the iron atoms in the haems, and this is associated with subtle changes in the three-dimensional conformation of the whole molecule.

The fact that Hb S differed in its electrophoretic properties from Hb A implied that the two molecules differed in structure, and it was soon shown that the difference did not lie in the haem groups but was present in the protein proper. The nature of the difference was elucidated by Ingram (1957, 1959) who found that a particular position in the aminoacid sequence of the β-polypeptide chain which is occupied by a glutamic acid residue in Hb A, is occupied by a valine residue in Hb S.

The method originally used to demonstrate this difference in the proteins is now generally referred to as 'finger printing' and it has been widely applied in the study of the primary structures of many different proteins and their genetically determined variants. The protein is first digested with a specific proteolytic enzyme such as trypsin or chymotrypsin, so that the polypeptide chains are split at a number of separate points to give a mixture of different smaller peptides. These peptides are then separated two-dimensionally on filter paper, first by electrophoresis and then by chromatography. The pattern of peptide 'spots' so produced is in general characteristic for the particular protein. In the case of haemoglobin, Ingram found that after digestion with trypsin which splits lysyl or arginyl peptide bonds, a complex pattern of peptides was obtained (fig. 1.2). The peptide pattern obtained

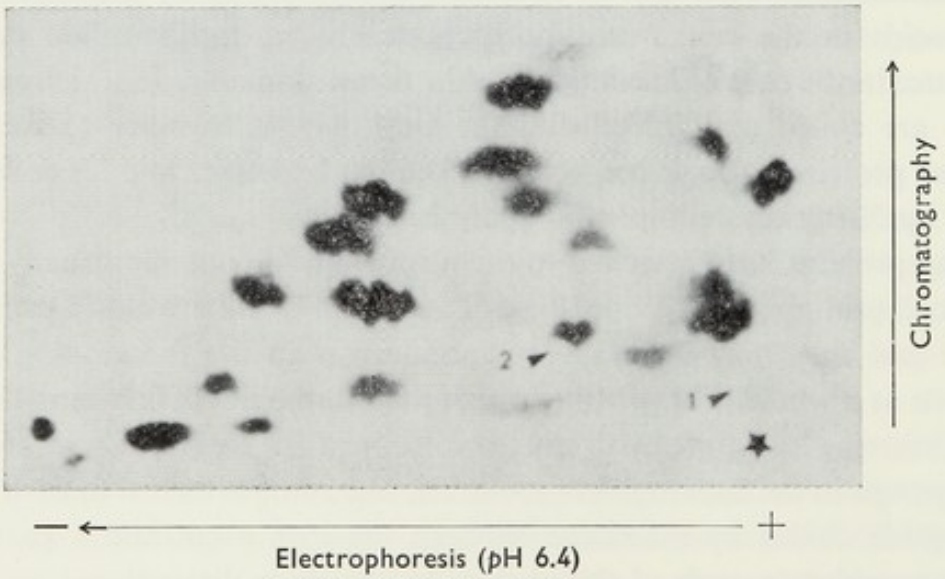


Fig. 1.2. 'Fingerprint' of Hb S. The peptides obtained by digestion of the haemoglobin have been separated on filter paper first by electrophoresis (horizontal dimension) and then by chromatography (vertical dimension). The peptide 'spots' were then developed with ninhydrin. *Key*: ★: point of application of peptide mixture to sheet of filter paper. **1**: position of peptide present in Hb A but not in Hb S. **2**: peptide present in Hb S but not in Hb A. (From Lehmann and Huntsman 1966.)

from Hb A appeared in most respects to be identical with that from Hb S. However there was one peptide present in Hb A which was not present in the Hb S pattern, and one peptide present in Hb S which was not seen in Hb A. These two peptides were isolated and the sequences of aminoacids in them determined. They each contained eight aminoacid residues, and their sequences were the same except that in the Hb A peptide the sixth residue from the amino end of the chain was glutamic acid, while in Hb S it was valine (fig. 1.3). Subsequently it was shown that all the other peptides

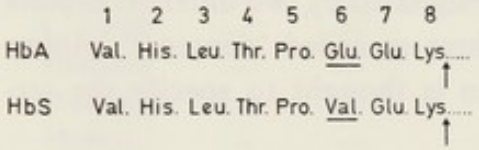


Fig. 1.3. Aminoacid sequences of the peptide present in tryptic digests of Hb A but not Hb S, and of the peptide present in tryptic digests of Hb S but not Hb A. The sequences represent the first eight aminoacids of the β-polypeptide chains of the two haemoglobins. They differ only at position 6. The arrows indicate the site of cleavage by trypsin. (For aminoacid abbreviations see caption fig. 1.5, p. 13.)

obtained from the two haemoglobins were the same as one another. It was also found that the peptides in the Hb A and Hb S which differed, represented the sequences at the amino-terminal end of the β-chains. Thus it could be concluded that the two haemoglobins differed only in a single aminoacid residue at the sixth position from the amino-terminal end of the β-chain, and that the sequence of the rest of the β-chain and the sequence of the whole of the α-chain was identical in the two molecules. Conventionally the aminoacid residues in a polypeptide chain are numbered from the amino-terminal end. If then the structure of Hb A is written as $\alpha_2\beta_2$, the structure of Hb S may be written as $\alpha_2\beta_2^{6\ \mathrm{Glu}\rightarrow\mathrm{Val}}$.

Hb S was the first example of a genetically determined protein variant in which the structural peculiarity was precisely identified. The result obtained was remarkable for its simplicity. A single gene difference such as that between the sickle-cell gene and its normal allele could be presumed to be the result of a single mutational step. This is the smallest unit of inherited variation. It could evidently lead to the smallest unit difference in the primary structure of a specific protein, namely the substitution of a single aminoacid residue by another. Subsequent work has shown that this is a general rule. Lehmann and Carrell (1969) list more than eighty other variant forms of haemoglobin which have now been shown to differ from the normal by a single aminoacid substitution, occurring at one position or another in either the α- or the β-chains (fig. 1.4). The same has also been found to be the

case for a variety of other altered proteins attributable to single gene mutations both in man and in other species.

1.4. Single aminoacid substitutions and the genetic code

The 64 base triplets which may occur in mRNA and which of course correspond to 64 complementary triplets in DNA, are shown in fig. 1.5, together with the various aminoacids which they are believed to specify in polypeptide synthesis. This is the so-called genetic code, and it is probably much the same for all species. Many aminoacids are designated by two different triplets and some by four or six. There are also three so-called 'nonsense' triplets whose position in the polynucleotide chain is thought to specify the termination of polypeptide chain synthesis.

It is instructive to examine the single aminoacid substitutions of the different haemoglobin variants in terms of these base triplets or codons. The mutational change from Hb A to Hb C for example has caused the replacement of a glutamic acid residue in the sixth position of the β-chain by a lysine residue. From fig. 1.5 one finds that the nucleotide triplet which codes for glutamic acid is either GAA or GAG, and the nucleotide triplet that codes for lysine is either AAA or AAG. So the mutation need have involved the change of only a single base in the triplet (i.e. GAA→AAA or GAG→AAG). All the other single aminoacid substitutions found in the abnormal haemoglobins may be considered in the same way. It turns out that in each case a single mutation need have involved no more than a single base change in a particular triplet. A similar conclusion is reached when mutations resulting in single aminoacid substitutions in other proteins and in other quite different species are considered. Taken as a whole the data which are now very extensive provide strong support for the view that most gene mutations represent simply a change of a single base in the whole base sequence of the DNA of a particular gene.

The normal β-polypeptide chain of human haemoglobin which contains 146 aminoacid residues, is presumably represented in the corresponding gene by a specific sequence of 438 bases (146 × 3). The mutation from the normal gene to the gene determining Hb C results in a single aminoacid substitution in the sixth position in the polypeptide sequence. Presumably therefore it occurred in the sixth base triplet in the corresponding stretch of DNA, and from the code we may infer that the mutation actually involved a change in the sixteenth base of the whole sequence. Similarly we can infer that the mutation from the normal gene to the sickle-cell gene involved the

α CHAIN VARIANTS

	1 Val…	5 Ala…	12 Ala..	15 Gly…	23 Glu…	30 Glu…	43 Phe…	47 Asp…	54 Gln…	58 His…	68 Asn…	80 Leu…	85 Asp.	87 His..	90 Lys.	92 Arg…	102 Ser…	112 His.	114 Pro…	136 Leu…	141 Arg
J. Toronto	—	Asp	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
J. Paris	—	—	Asp	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
J. Oxford	—	—	—	Asp	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Memphis	—	—	—	—	Gln	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
G. Audhali	—	—	—	—	Val	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Chad	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
G. Chinese	—	—	—	—	—	Gln	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Torino	—	—	—	—	—	—	Val**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
L. Ferrara	—	—	—	—	—	—	—	Gly	—	—	—	—	—	—	—	—	—	—	—	—	—
Hasharon	—	—	—	—	—	—	—	His	—	—	—	—	—	—	—	—	—	—	—	—	—
Shimonoseki	—	—	—	—	—	—	—	—	Arg	—	—	—	—	—	—	—	—	—	—	—	—
Mexico	—	—	—	—	—	—	—	—	Glu	—	—	—	—	—	—	—	—	—	—	—	—
M. Boston	—	—	—	—	—	—	—	—	—	Tyr*	—	—	—	—	—	—	—	—	—	—	—
G. Philadelphia	—	—	—	—	—	—	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—	—
Ube II	—	—	—	—	—	—	—	—	—	—	Asp	—	—	—	—	—	—	—	—	—	—
Ann Arbor	—	—	—	—	—	—	—	—	—	—	—	Arg	—	—	—	—	—	—	—	—	—
G. Norfolk	—	—	—	—	—	—	—	—	—	—	—	—	Asn	—	—	—	—	—	—	—	—
M. Iwate	—	—	—	—	—	—	—	—	—	—	—	—	—	Tyr*	—	—	—	—	—	—	—
Broussais	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Asn	—	—	—	—	—	—
J. Cape Town	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Gln	—	—	—	—	—
Chesapeake	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Leu	—	—	—	—	—
Manitoba	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Arg	—	—	—	—
Dakar	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Gln	—	—	—
Chiapas	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Arg	—	—
Bibba	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Pro**	—
Singapore	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Pro

Fig. 1.4a. Aminoacid substitutions in 26 different α-chain variants of human haemoglobin. For aminoacid abbreviations see caption fig. 1.5 p. 13.

Key:
* methaemoglobinaemia
** unstable haemoglobins

β CHAIN VARIANTS

	1 Val...	6 Glu	7 Glu...	16 Gly...	22 Glu...	26 Glu.	28 Leu...	42 Phe	43 Glu.	46 Gly...	61 Lys.	63 His...	67 Val...	77 His.	79 Asp...	87 Thr..	90 Glu.	92 His..	95 Lys..	98 Val...	113 Val...	121 Glu...	126 Val...	132 Lys...	136 Gly...	146 His
S	—	Val	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
C	—	Lys	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
G. San José	—	—	Gly	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Siriraj	—	—	Lys	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
J. Baltimore	—	—	—	Asp	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
D. Bushman	—	—	—	Arg	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
E. Saskatoon	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
G. Coushatta	—	—	—	—	Ala	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
E.	—	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Genova	—	—	—	—	—	—	Pro**	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Hammersmith	—	—	—	—	—	—	—	Ser**	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
G. Galverston	—	—	—	—	—	—	—	—	Ala	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
K. Ibadan	—	—	—	—	—	—	—	—	—	Glu	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Hikari	—	—	—	—	—	—	—	—	—	—	Asn	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
M. Saskatoon	—	—	—	—	—	—	—	—	—	—	—	Tyr*	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Zurich	—	—	—	—	—	—	—	—	—	—	—	Arg**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
M. Milwaukee	—	—	—	—	—	—	—	—	—	—	—	—	Glu*	—	—	—	—	—	—	—	—	—	—	—	—	—
Sidney	—	—	—	—	—	—	—	—	—	—	—	—	Ala**	—	—	—	—	—	—	—	—	—	—	—	—	—
J. Iran	—	—	—	—	—	—	—	—	—	—	—	—	—	Asp	—	—	—	—	—	—	—	—	—	—	—	—
G. Accra	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Asn	—	—	—	—	—	—	—	—	—	—	—
D. Ibadan	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—	—
Agenogi	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Lys	—	—	—	—	—	—	—	—	—
M. Hyde Park	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Tyr*	—	—	—	—	—	—	—	—
N. Baltimore	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Glu	—	—	—	—	—	—	—
Köln	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Met**	—	—	—	—	—	—
New York	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Glu	—	—	—	—	—
D. Punjab	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Gln	—	—	—	—
O. Arab	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Lys	—	—	—	—
Hofu	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Glu	—	—	—
K. Woolwich	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Gln	—	—
Hope	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	Asp	—

Fig. 1.4b. Aminoacid substitutions in 31 different β-chain variants of human haemoglobin. Detailed references and further examples of both α and β chain variants are given by Lehmann and Carrell (1969).

Key: ▭ Sickle cell haemoglobin
* methaemoglobinaemias
** unstable haemoglobins

adjacent base in the same triplet (GAA→GUA or GAG→GUG). This would represent the seventeenth base in the full sequence. Similar precise derivations may be made for most of the other variant haemoglobins where the nature of the aminoacid substitution has been determined.

It is convenient (Crick 1967) to display the various single aminoacid substitutions detected in mutants of a given protein by an arrow on the diagram of the genetic code as shown for the haemoglobin variants in fig. 1.6. If the substitution is due to single base change in the first position of the triplet, the arrow marking the change will be vertical and will begin and end in the same relative position within a square on the figure. If the change is in

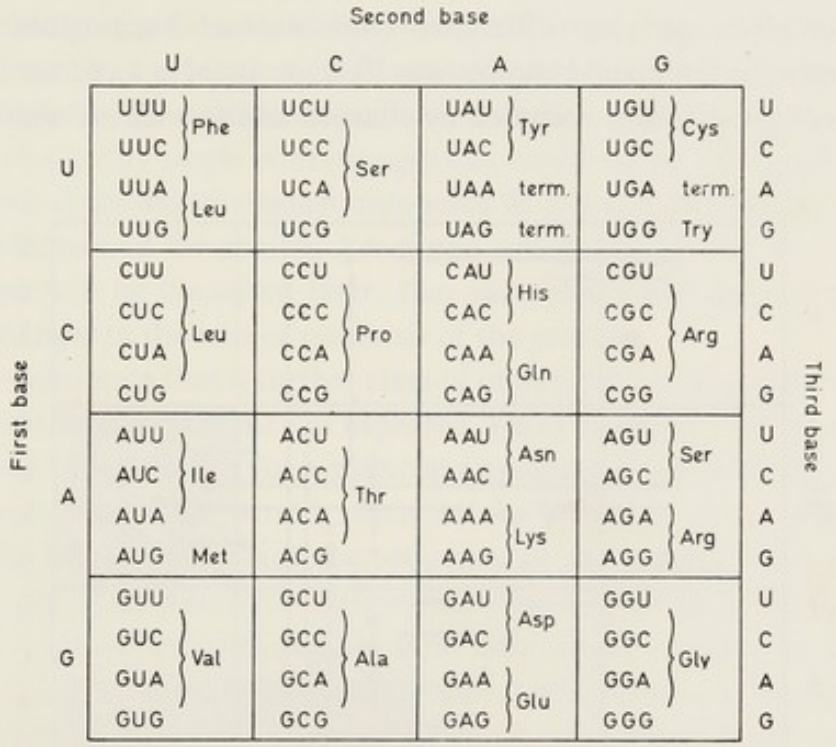


Fig. 1.5. The genetic code. [For detailed references to derivation of the code see Cold Spring Harbor Symp. Quant. Biol. *31* (1966), Crick (1967), Woese (1967) and Yčas (1969).]
Bases: U: uracil, C: cytosine, A: adenine, G: guanine.
Aminoacids:

Ala:	alanine	Gly:	glycine	Pro:	proline
Arg:	arginine	His:	histidine	Ser:	serine
Asn:	asparagine	Ile:	isoleucine	Thr:	threonine
Asp:	aspartic acid	Leu:	leucine	Try:	tryptophan
Cys:	cysteine	Lys:	lysine	Tyr:	tyrosine
Gln:	glutamine	Met:	methionine	Val:	valine
Glu:	glutamic acid	Phe:	phenylalanine		

term. = 'nonsense' triplet – chain termination.

the second base, the arrow will be horizontal. If the change is in the third base, the arrow will be vertical but will begin and end in the same square. Thus if an aminoacid substitution requires at a minimum only a single base change in a coding triplet then the arrow will be either horizontal or vertical, and this is seen to be so in all cases. A diagonal arrow would imply that at least two bases had been changed, and this would have been expected in a considerable fraction of the cases if the changes in aminoacids were quite arbitrary and at random.

It will be seen that the majority of the arrows are localised to particular areas of the diagram. This arises because of the method by which most of the variant haemoglobins were discovered. They were detected because they showed an electrophoretic difference from normal haemoglobin A, and consequently the aminoacid change was likely to involve a change in charge. The areas of the diagram occupied by charged aminoacids are shaded, while

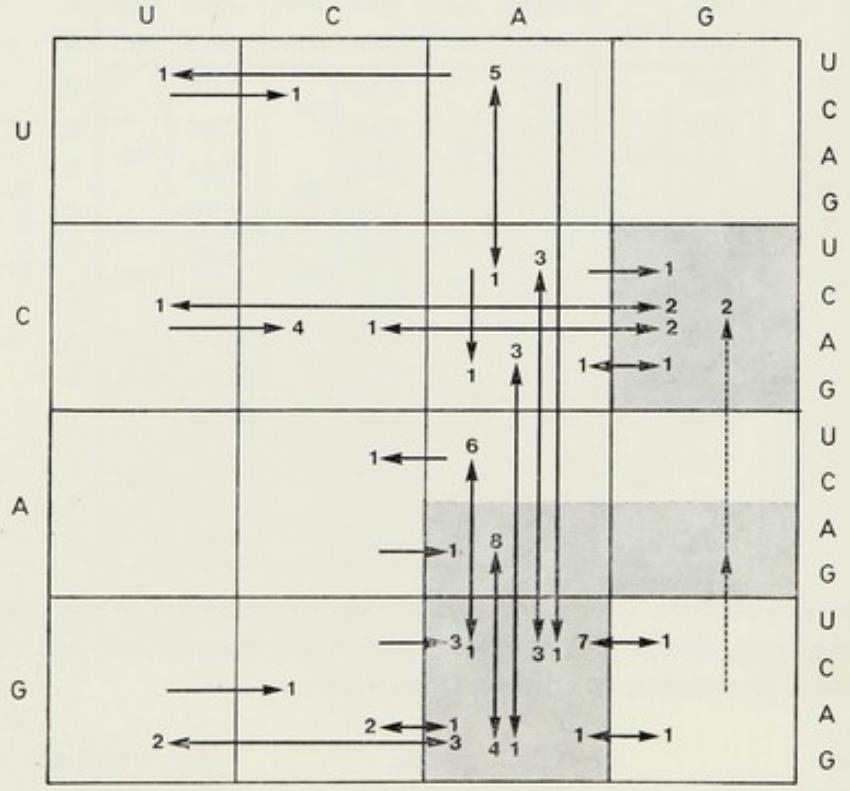


Fig. 1.6. Diagram of single aminoacid substitutions in 79 different α- or β-chain haemoglobin variants (after Crick 1967). The code is as arranged in fig. 1.5. Each arrow represents a single aminoacid substitution, and the number attached to an arrow gives the number of different variants with that particular aminoacid substitution. The shaded areas of the diagram represent charged aminoacids (i.e. the basic aminoacids lysine and arginine *or* the acidic aminoacids aspartic acid and glutamic acid). The haemoglobin variants shown are those listed by Lehmann and Carrell (1969).

the remainder is occupied by neutral aminoacids. As expected from the mode of discovery a large proportion of the arrows start or end in a shaded region. But the widespread use of electrophoresis in the search for variants of haemoglobin has largely been a matter of technical convenience. One may therefore anticipate that there exist many other variants of haemoglobin not involving a change in charge, and therefore remaining undetected. The arrows representing such mutations would presumably occur in the unfilled areas of the diagram shown in fig. 1.6. No doubt these 'concealed' mutations will be progressively uncovered as other techniques are more widely applied in the search for variants.

Furthermore, not all mutations involving a single base change in a gene determining a particular polypeptide chain will result in an altered or variant protein. This is because the code is degenerate and the majority of aminoacids are represented by two different nucleotide triplets, and some by four or six. It follows that many single base changes can occur which will result in a new triplet which codes for the same aminoacid as the original triplet. Thus the primary structure of the protein concerned will not be altered. It is possible, however, as will be discussed later, that such mutations may on occasion result in changes in the rate of synthesis of the protein.

One may also note that a further class of mutations having quite a different effect on protein structure may be expected. These would involve base changes which alter a triplet coding for a particular aminoacid to one which codes for chain termination. This would result in the synthesis of an abbreviated polypeptide which would lack the sequence of aminoacids from the point affected by the mutation to the carboxyl terminal end of the normal polypeptide. Such a shortened polypeptide is in most cases unlikely to assume a stable three-dimensional conformation and would in general be expected to differ profoundly in properties from its normal counterpart. Such mutations, although they have been demonstrated under rather special circumstances in certain microorganisms, are likely to be very difficult to identify with certainty in man and may perhaps in most instances be manifest simply as a protein or enzyme deficiency.

Clearly a very considerable number of different single base alterations to the DNA sequence of a gene coding for a particular polypeptide chain could arise as a result of separate mutations. Their effects on the structure of the polypeptide will vary according to the position in the sequence at which they occur and the specific nature of the base alteration. A general picture of the relative frequencies of the different kinds of effects that may ensue can be obtained by considering the sequence of 438 bases which code for 146

aminoacids in the β-polypeptide chain of haemoglobin. Of all the possible single base changes that could occur, about 25% would cause no alteration in the aminoacid sequence, about 2–3% would result in shortening of the polypeptide by premature chain termination, and about 73% would be expected to result in a single aminoacid substitution in the polypeptide. Of those that result in single aminoacid substitutions, only about one-third would be expected to cause an alteration in charge which might allow the variant protein to be detected electrophoretically.

1.5. The effects of single aminoacid substitutions

We have seen that in any particular protein separate mutations can result in a considerable number of different single aminoacid substitutions. Some of these may cause little or no change in the structural integrity of the protein molecule or in its other properties, so that the effects of the mutation on function may be very slight and perhaps undetectable. In other cases however the protein may be altered in such a way as to give rise to a variety of secondary pathological and clinical consequences.

The specific change in the properties of a protein produced by a particular aminoacid substitution will of course depend on the chemical nature of the aminoacid substituted. Whether for example the substitution involves a change from a hydrophobic to a hydrophilic residue, a change in ionisation, or a marked change in the physical dimensions of the side chain. Furthermore it will depend on the particular aminoacid site in the three-dimensional structure of the protein which is substituted. In the case of haemoglobin it appears that substitutions occurring at sites on the external surface of the molecule are in general less likely to result in significant effects than changes elsewhere. The most critical sites appear to be those which involve areas of contact between the different polypeptide chains in the tetramers or between the polypeptide chains and the haem groups, and also other internal sites which happen in one way or another to be critical for maintaining the conformational integrity of the molecule (Perutz and Lehmann 1968).

An alteration in protein structure of this sort is probably the specific underlying biochemical cause of many different forms of inherited disease. The characteristic features of such a disease will be determined by the nature and function of the protein involved, and the manner in which its properties have been changed. In such conditions therefore it should eventually be possible to trace in detail the precise way in which the alteration in the primary structure of the protein leads to functional changes in the molecule, and how

these in turn give rise to the pathological disturbances that are manifest. So far however a satisfactory insight into this causal sequence has been achieved in relatively few conditions. However the general nature of the problem is well illustrated by the various and in many cases quite strikingly different disorders due to structural abnormalities in haemoglobin. They serve as models for the kind of analysis of inherited diseases which should become increasingly possible as the structures of other proteins (particularly enzyme proteins) and of their variants are elucidated.

1.5.1. Sickle-cell disease

Hb S was first distinguished from Hb A because of the difference in electrophoretic mobility. Shortly afterwards however another difference but in a quite different property was discovered. It was found (Perutz and Mitchison 1950) that deoxygenated sickle-cell haemoglobin is very much less soluble than deoxygenated normal haemoglobin, whereas in the oxygenated state

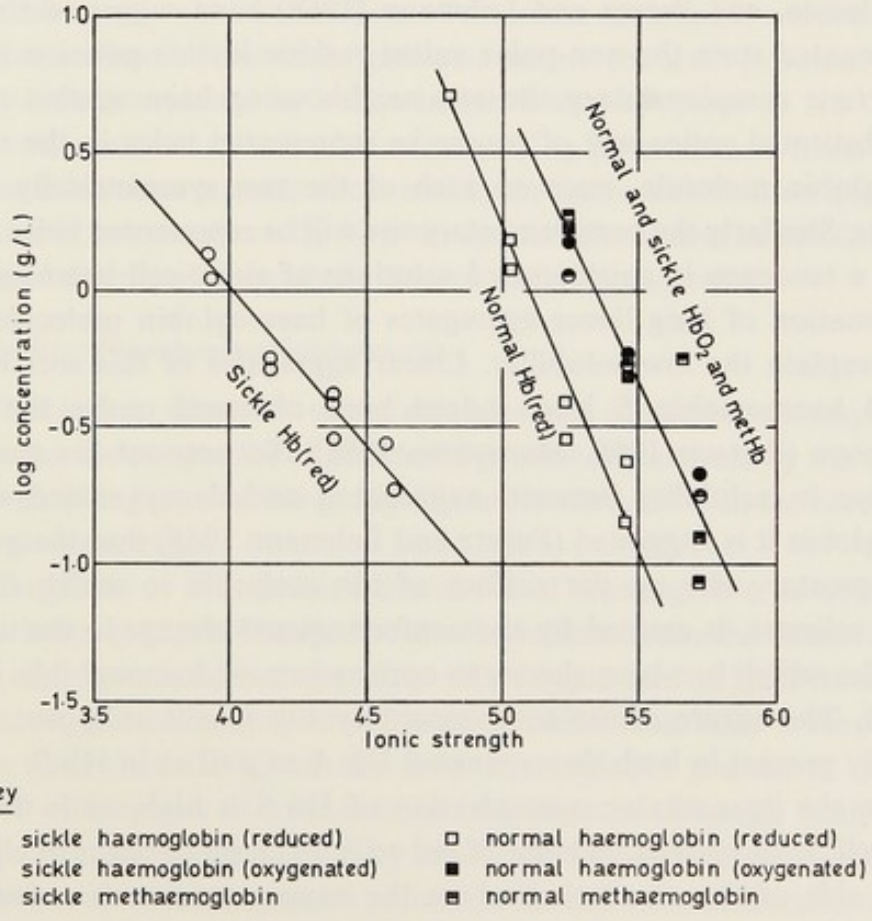


Fig. 1.7. Solubilities of normal and sickle-cell haemoglobins plotted against ionic strength of solution. (From Perutz and Mitchison 1950.)

the two haemoglobins are equally soluble (fig. 1.7). This finding immediately suggested a simple explanation for the sickling phenomenon which occurs when red cells containing Hb S are subjected to low oxygen tension. When deoxygenated, Hb S which is present in high concentrations in the red cell would tend to come out of free solution and cause a characteristic deformation of the cell shape. It can indeed be shown that concentrated sickle-cell haemoglobin solutions free of red cell stroma become increasingly viscous as the oxygen tension is reduced and eventually assume a semi-solid gel-like state, in which spindle-shaped birefringent bodies, 1–15 μ in length, can be seen microscopically (Harris 1950). These are remarkably similar in shape to the elongated sickle-like forms that intact red cells containing Hb S take up at low oxygen tension.

The aminoacid substitution in sickle-cell haemoglobin involves a replacement of a glutamic acid in haemoglobin A by a valine, i.e. the substitution of a polar by a non-polar residue. The site at which this occurs, the sixth position in the β-polypeptide chain, has been shown to lie at the surface of the molecule, and Perutz and Lehmann (1968) have suggested that in the deoxygenated state the non-polar valine residue in this position is able to adhere to a complementary site on a neighbouring haemoglobin molecule. The substituted valine will of course be represented twice in the sickle-cell haemoglobin molecule, once on each of the two symmetrically arranged β-chains. Similarly the complementary site will be represented twice. So there will be a tendency in concentrated solutions of sickle-cell haemoglobin for the formation of long linear aggregates of haemoglobin molecules, which would explain the low solubility. Linear aggregates of this sort in deoxygenated haemoglobin S have indeed been observed under the electron microscope (Stetson 1966, Murayama 1966). To account for the marked difference in solubility between oxygenated and deoxygenated sickle-cell haemoglobin it is suggested (Perutz and Lehmann 1968) that the postulated complementary site on the surface of the molecule to which the valine residue adheres, is created by the conformational change in the tetrameric molecules which has been shown to occur when oxyhaemoglobin is deoxygenated. The nature of the complementary site is still unknown, but it is probably present in both deoxygenated Hb A as well as in Hb S.

When the intracellular concentration of Hb S is high, as is the case in sickle-cell homozygotes, sickling of red cells will tend to occur *in vivo* on the venous side of the circulation where the oxygen tension is reduced. This results in increased viscosity of the blood and this is particularly likely to impede the circulation in the smaller veins, and the venous side of the

capillaries. This will tend to cause further deoxygenation and increased sickling and so a vicious circle is likely to be set up accentuating the effect. The sickled cells may also block the smaller blood vessels by forming thrombi, and these probably lead to the multiple scattered foci of tissue destruction which are a characteristic feature of the disease and which can result in a wide variety of symptoms. The deformed red cells also tend to be broken down at a much increased rate. Thus one can see at least in a general way, how both the haemolytic anaemia and the other diverse pathological manifestations which appear in the disease may be traced back to the alteration in the solubility of the haemoglobin produced by the single specific aminoacid substitution.

It is of interest to contrast sickle-cell anaemia with haemoglobin C disease. Sickle-cell anaemia is a severe disorder which is often fatal in childhood or adolescence. Haemoglobin C disease is by comparison benign. The degree of anaemia that occurs is relatively slight and in many cases the affected individuals live a normal and active life with no obvious disability. Yet the two conditions are due to the substitution of the same aminoacid residue in the haemoglobin molecule. This is the glutamic acid in the sixth position on the β-chains, which is replaced by valine in Hb S and by lysine in Hb C. But while the valine substitution leads to the sickling phenomenon from which most of the deleterious consequences ensue, the lysine substitution does not. Indeed it is still not clear exactly how the lysine substitution in Hb C causes even the mild degree of haemolytic disease which is observed.

1.5.2. Hereditary methaemoglobinaemias

Another illustration of how the site of an aminoacid substitution is important in determining the properties of an altered protein is provided by the abnormal haemoglobins found in the group of conditions known as the hereditary methaemoglobinaemias.

Methaemoglobin is the oxidised derivative of haemoglobin in which the iron of the haem group is changed to the ferric state from the usual ferrous state. It is unable to combine reversibly with oxygen and so fulfil its normal function in oxygen transport. In the normal individuals only a very small proportion (less than 0.5%) of the total haemoglobin present in the circulating red cells occurs as methaemoglobin, because although methaemoglobin is constantly being formed by oxidation of haemoglobin, there are powerful enzymic reducing systems present in the red cell which reconvert the haem iron back to the ferrous state. The result is an equilibrium which may be written:

$$\text{Hb } (Fe_4^{++}) \rightleftarrows \text{Hb } (Fe_3^{++}Fe^{+++}) \rightleftarrows \text{Hb } (Fe_2^{++}Fe_2^{+++}) \rightleftarrows \text{Hb } (Fe^{++}Fe_3^{+++} \rightleftarrows \text{Hb } (Fe_4^{+++})$$

Haemoglobin — Methaemoglobin

and which is normally kept well over to the left. A number of rare inherited abnormalities are known however in which a markedly increased proportion of the haem iron is in the ferric state. These disorders are known as hereditary methaemoglobinaemias and affected individuals are characterised by the blue cyanotic appearance which occurs when a significant proportion of the haemoglobin in the circulating red cells is not combined with oxygen.

Some cases of hereditary methaemoglobinaemia are due to a genetically determined defect of an enzyme (methaemoglobin reductase, see p. 254) which is normally concerned with maintaining haem iron in the ferrous state. Other cases are due to the presence of an abnormal haemoglobin in which the ferric state is unusually stable and so is not readily reduced by the normal enzyme systems. Several different haemoglobin variants are known in which this effect occurs and they each involve single aminoacid substitutions in the polypeptide chains in the immediate neighbourhood of the affected haems.

The four haem groups of haemoglobin lie in separate pockets on the surface of the haemoglobin molecule formed by folds in each of the four polypeptide chains. Each haem is attached to its polypeptide chain by a co-ordinate linkage from the iron to a specific histidine residue (fig. 1.8). This is at position 87 in the α-chain and position 92 in the β-chain. The iron is also linked through an oxygen molecule in oxygenated haemoglobin or a water molecule in reduced haemoglobin to another histidine residue on the opposite side of the fold (position 58 in the α-chain and position 63 in the β-chain).

One of the variant haemoglobins which gives rise to methaemoglobinaemia is known as Hb M Boston, and it has a substititution at position 58 in the α-chain where the normal histidine residue is replaced by a tyrosine residue (Gerald and Efron 1961). This position is spatially very close to the α-chain haem iron. When this haem iron in a given molecule is in the ferrous state nothing probably happens, but when, as will occur sooner or later, it becomes oxidised to the ferric state, the phenol side chain of the substituted tyrosine residue at position 58 is in just the right position to bond to the ferric iron. This forms a very stable complex which is not readily reduced by the normal methaemoglobin reductase system present in the red cell. Consequently the irons of the haem groups associated with the two α-chains of the molecule remain in the ferric state and are incapable of combining with and therefore

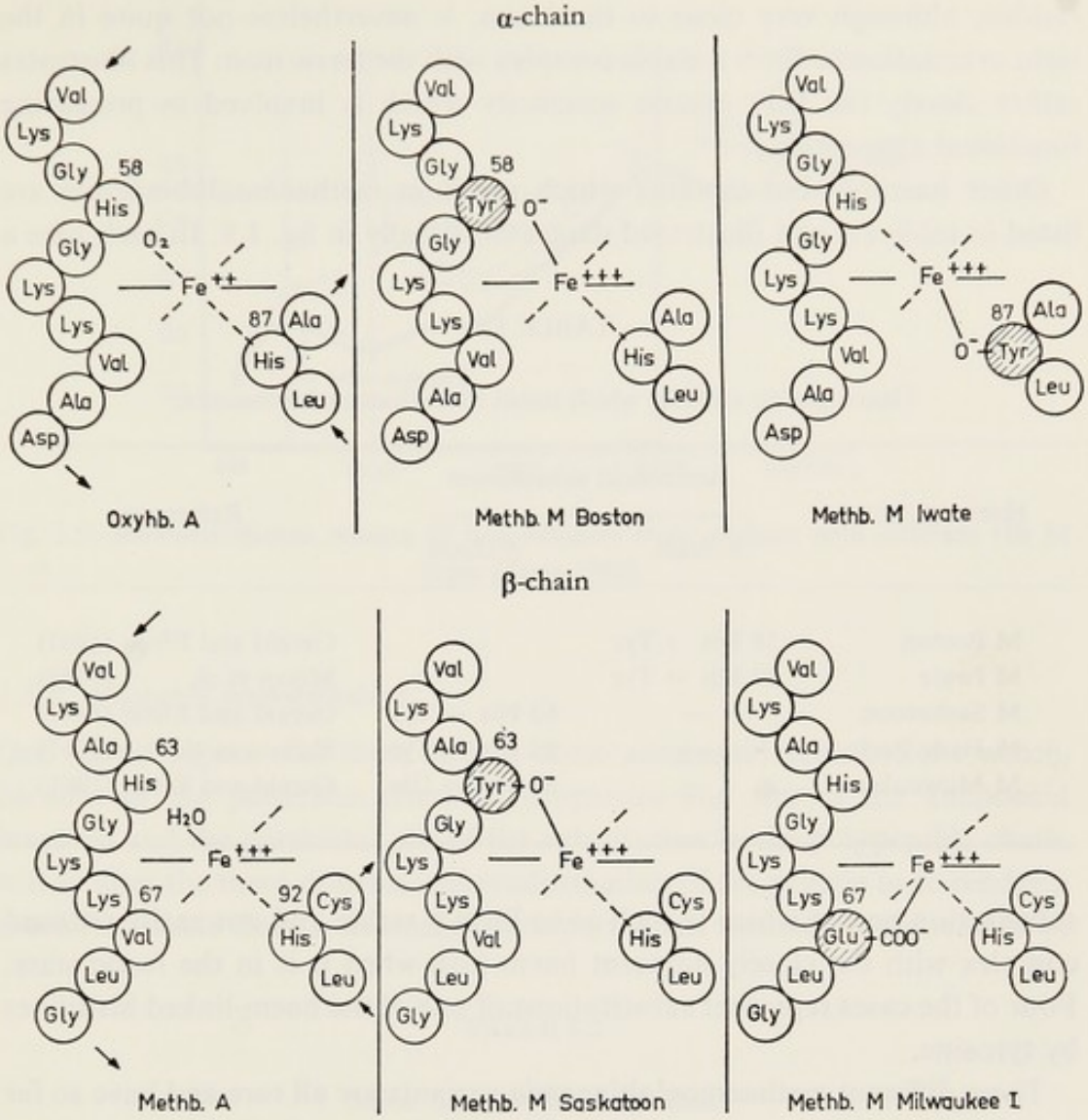


Fig. 1.8. Diagram showing effects of aminoacid substitutions in different types of Hb M. The substitution in each case involves a position close to the haem iron, so that an internal complex can be formed between the iron in the ferric state, and the phenolic side chain of a substituted tyrosine or the negatively charged side chain of glutamic acid (Tönz 1968).

transporting oxygen. The haems of the two β-chains are however not directly affected.

It is interesting that another haemoglobin variant is known, Hb Norfolk (Ager et al. 1958), in which the aminoacid substitution, a replacement of a glycine residue by an aspartic acid residue, occurs at position 57 of the α-chain (Baglioni 1962b), immediately adjacent to the position affected in Hb M Boston. In this case, however, methaemoglobinaemia does not occur. This is presumably because the carboxyl side chain of the aspartic acid

residue, although very close to the haem, is nevertheless not quite in the right orientation to form a stable complex with the ferric iron. This illustrates rather clearly the very precise specificity which is involved in producing functional abnormality.

Other haemoglobin variants which result in methaemoglobinaemia are listed in table 1.1 and illustrated diagrammatically in fig. 1.8. In each case a

TABLE 1.1

Haemoglobin variants which result in methaemoglobinaemia.

Haemoglobin	Aminoacid substitution		Reference
	α-chain	β-chain	
M Boston	58 His → Tyr	—	Gerald and Efron (1961)
M Iwate	87 His → Tyr	—	Miyaji et al. (1963)
M Saskatoon	—	63 His → Tyr	Gerald and Efron (1961)
M Hyde Park	—	92 His → Tyr	Shibata et al. (1967)
M Milwaukee	—	67 Val → Glu	Gerald and Efron (1961)

substitution occurs which is such as to form a stable and not easily reduced complex with the closely adjacent haem iron when it is in the ferric state. Four of the cases represent substitutions of one of the haem-linked histidines by tyrosine.

These different methaemoglobinaemia variants are all rare and have so far only been seen in heterozygotes. In affected individuals both the normal and the abnormal haemoglobin are present in the red cells, and in the abnormal haemoglobin only either the haems of the α-chains or the haems of the β-chains are kept in the ferric state. This is nevertheless sufficient to produce marked methaemoglobinaemia and obvious cyanosis.

One interesting indication of the intramolecular changes that occur in these abnormal methaemoglobins is a characteristic alteration in their absorption spectra (Gerald and Scott 1966). The spectral difference involves a shift of the normal absorption maximum at 632 mμ to a lower wavelength (around 600 mμ). The extent of this shift, however, varies in the different abnormal methaemoglobins (fig. 1.9) and presumably depends on the characteristic distortion of the three-dimensional structure resulting from the different aminoacid substitutions.

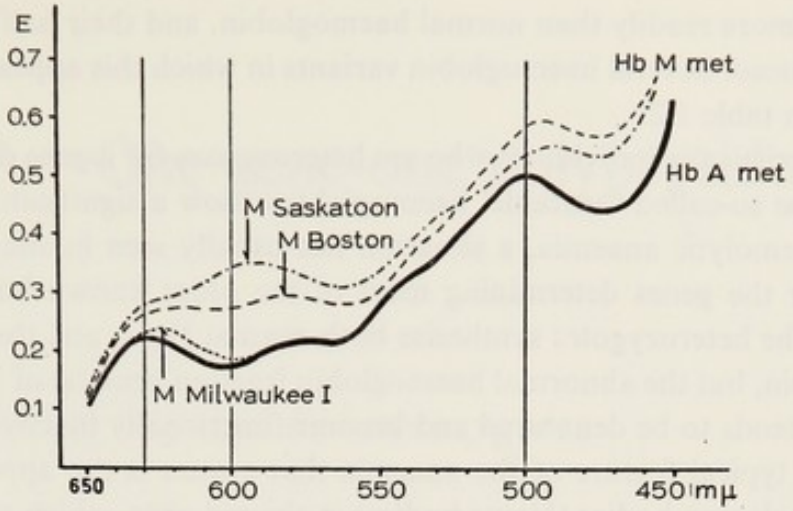


Fig. 1.9. Methaemoglobin spectra of haemolysates from patients with different Hb M types (Tönz 1968).

1.5.3. Unstable haemoglobins

One would expect that there will be some aminoacid substitutions which, because of the particular chemical properties and size of the aminoacid involved and the particular site of the substitution in the polypeptide chain, will so alter the three-dimensional conformation of the protein as to render it much less stable than its normal counterpart. Such variants would tend to be

TABLE 1.2

Some 'unstable' haemoglobins.

Haemoglobin	Aminoacid substitution		References
	α-chain	β-chain	
Torino	43 Phe → Val	—	Beretta et al. (1968)
Bibba	136 Leu → Pro	—	Kleihauer et al. (1968)
Genova	—	28 Leu → Pro	Sansone et al. (1967)
Hammersmith	—	42 Phe → Ser	Dacie et al. (1967)
Zürich	—	63 His → Arg	Muller and Kingma (1961)
Sydney	—	67 Val → Ala	Carrell et al. (1966)
Santa Ana	—	88 Leu → Pro	Opfell et al. (1968)
Borås	—	88 Leu → Arg	Hollender et al. (1969)
Köln	—	98 Val → Met	Carrell et al. (1966)

denatured more readily than normal haemoglobin, and their half-life *in vivo* greatly reduced. Several haemoglobin variants in which this appears to occur are listed in table 1.2.

Characteristically individuals who are heterozygous for a gene determining one of these so-called 'unstable' haemoglobins show a significant degree of chronic haemolytic anaemia, a situation not usually seen in simple heterozygotes for the genes determining most of the other known haemoglobin variants. The heterozygotes synthesise both normal Hb A and the abnormal haemoglobin, but the abnormal haemoglobin fraction because of its inherent instability tends to be denatured and become functionally inactive relatively rapidly. A typical feature of the anaemia that ensues is the appearance of so-called inclusion bodies (Heinz bodies) in the red cells, which are particularly prominent in patients who have been splenectomised. These appear to be composed of denatured haemoglobin, and they apparently render the red cells abnormally susceptible to premature destruction in the spleen.

The 'unstable' haemoglobin variants listed in table 1.2 mainly involve substitutions at internal sites in the molecular configuration which are normally occupied by non-polar aminoacids. The substitution in most of the cases is to another non-polar aminoacid with slightly different molecular dimensions. The apparent paucity of examples of substitutions to polar aminoacids with ionised side-chains is significant. It probably arises because in order to detect and characterise an 'unstable' haemoglobin at all, it is necessary that the degree of instability should not be so severe that it is quite incompatible with the survival of the molecule. As Lehmann and Carrell (1969) have pointed out, replacement of a non-polar aminoacid at one of these internal sites in the molecule by a residue with an ionised side-chain would usually result in an almost if not totally non-viable molecule, and it would be unlikely for this to be detected and characterised by the methods currently available. So the anaemias due to the known 'unstable' haemoglobins, although usually in themselves quite severe, probably represent only the milder forms of this type of abnormality.

The manner in which the various known substitutions of this sort tend to distort the three-dimensional structure of the haemoglobin and so lead to molecular instability is discussed in detail by Perutz and Lehmann (1968).

CHAPTER 2

One gene-one polypeptide chain

2.1. *'Hybrid' proteins in heterozygotes*

2.1.1. Haemoglobin

When haemolysates from individuals with the sickle-cell trait are examined electrophoretically, the two characteristic forms of haemoglobin seen are Hb A, $\alpha_2\beta_2$, and Hb S, $\alpha_2\beta_2^{6\ Glu\rightarrow Val}$ (conveniently written $\alpha_2\beta_2^S$). One does not observe the other possible kind of haemoglobin $\alpha_2\beta\beta^S$, which might be expected to occur in individuals who are synthesising both the β and β^S polypeptide chains. This curious and for a long time very puzzling situation also obtains in heterozygotes for other alleles determining haemoglobin variants. For example in the haemoglobin C trait one finds Hb A, $\alpha_2\beta_2$, and Hb C, $\alpha_2\beta_2^C$, but not $\alpha_2\beta\beta^C$, and in sickle-cell–haemoglobin C disease one finds Hb S, $\alpha_2\beta_2^S$, and Hb C, $\alpha_2\beta_2^C$, but not $\alpha_2\beta^S\beta^C$.

It has in fact been shown (Guidotti et al. 1963, Benesch et al. 1966) that hybrid molecules of the general type $\alpha_2\beta\beta^S$ in sickle-cell trait, and $\alpha_2\beta^S\beta^C$ in sickle-cell–haemoglobin C disease, are almost certainly present in quantity in the red cells of such heterozygotes. They are not however usually detected because the identification of the individual protein components in a mixture of haemoglobins requires in the ordinary way their physical separation by methods such as electrophoresis or column chromatography, and also because haemoglobin very readily dissociates into half molecules.

It seems that in solution the haemoglobin tetramer $\alpha_2\beta_2$ partially dissociates to give dimers of structure $\alpha\beta$, and there is a rapid dissociation–association equilibrium which may be written $\alpha_2\beta_2 \rightleftarrows 2\alpha\beta$. Similarly for Hb S one has $\alpha_2\beta_2^S \rightleftarrows 2\alpha\beta^S$. In a solution containing both $\alpha_2\beta_2$ and $\alpha_2\beta_2^S$, the dimers $\alpha\beta$ and $\alpha\beta^S$ are of course present and there is also the third form of the tetramer $\alpha_2\beta\beta^S$. The two dimers and the three tetramers are in equilibrium and the whole situation may be represented by the scheme shown in fig. 2.1. When such a solution is subjected to electrophoresis or chromatography this

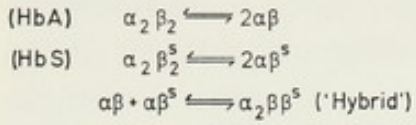


Fig. 2.1. Equilibria of haemoglobin tetramers and dimers in red cells of sickle-cell heterozygote (sickle-cell trait).

equilibrium is immediately disturbed by the preferential removal of $\alpha_2\beta_2$ and $\alpha_2\beta_2^S$ at the two extreme ends of the separation. As the most rapidly moving tetramer, say $\alpha_2\beta_2$, separates out of the mixture this effectively removes $\alpha\beta$ dimers and leads in turn to further dissociation of $\alpha_2\beta\beta^S$. Thus as the separation proceeds the amount of $\alpha_2\beta\beta^S$ is progressively reduced and by the time there is a complete separation of $\alpha_2\beta_2$ from $\alpha_2\beta_2^S$ no molecules of $\alpha_2\beta\beta^S$ remain (fig. 2.2). In consequence when haemolysates from sickle-cell

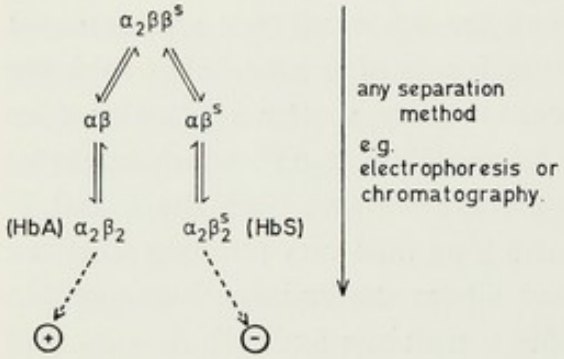


Fig. 2.2. Diagram showing how separation of haemoglobins in sickle-cell heterozygote haemolysate leads to the disappearance of the hybrid form, $\alpha_2\beta\beta^S$ (Benesch et al. 1966).

trait individuals are examined by these standard procedures only Hb A and Hb S are seen. The same general phenomenon occurs when heterozygotes for other haemoglobin alleles are studied.

The formation of 'hybrid' proteins in heterozygotes, that is the occurrence of a single protein molecule containing each of the structurally different polypeptide chains coded by the two alleles, appears to be a common and widespread phenomenon, and a number of examples involving a variety of different proteins and enzymes have now been identified. The phenomenon is of particular interest, because by its very nature a 'hybrid' molecule of this sort cannot be present in individuals homozygous for either allele. It therefore represents a special molecular form peculiar to the heterozygous state. Also since a heterozygote receives one allele from one parent and one from the other, this particular molecular form can occur in individuals neither of whose parents possess it.

It is the facility with which haemoglobin dissociates into half molecules which renders the 'hybrid' forms undetectable in this case by ordinary separation procedures. However many proteins which contain multiple

polypeptide chains are much less readily dissociated into subunits, and consequently such 'hybrid' forms that may exist in heterozygotes are more easily demonstrable.

2.1.2. Peptidase A

An example is provided by the enzyme known as peptidase A (Lewis and Harris 1967), which occurs in red blood cells as well as in most other tissues. It is a dipeptidase which has been shown to hydrolyse a considerable number of different dipeptides, and it can be demonstrated by a specific staining technique after electrophoresis.

In the course of electrophoretic studies of the enzyme in red cells from many different individuals a number of genetically determined variant types have been identified (Lewis and Harris 1967, Lewis et al. 1968). The type of pattern seen in the majority of people is called Pep A 1. The other types are in most populations relatively rare, but among Negroes two of them Pep A 2-1 and Pep A 2 occur with an appreciable frequency. Thus in many African populations 15% to 20% of people may be Pep A 2-1 and nearly 1% Pep A2.

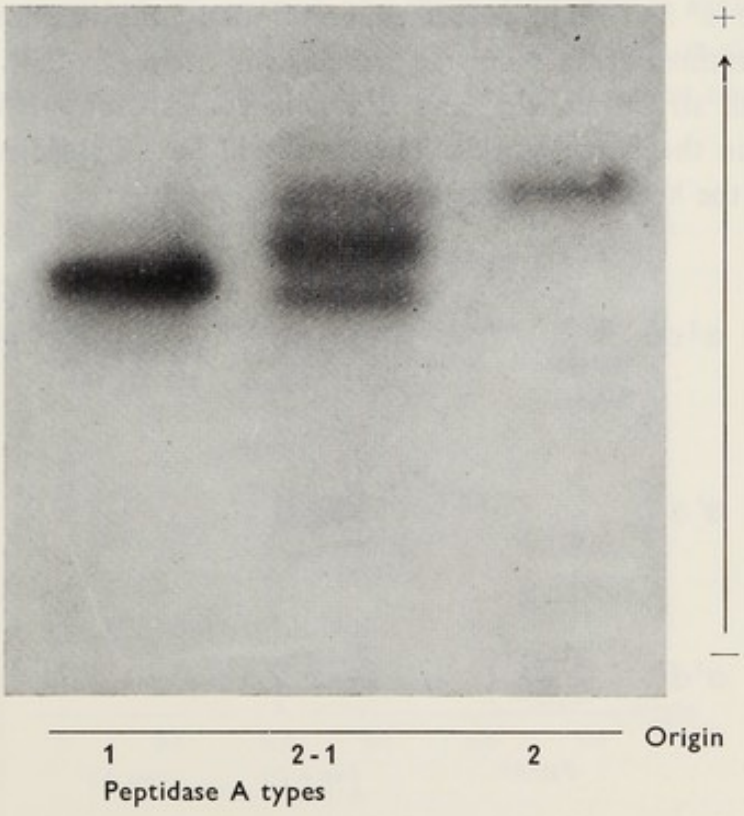


Fig. 2.3. Electrophoretic patterns of peptidase A types 1, 2-1, and 2. Electrophoresis and staining carried out as described by Lewis and Harris (1967).

Family studies show that the three peptidase types Pep A 1, Pep A 2-1 and Pep A 2 are determined by two autosomal genes (*Pep* A^1 and *Pep* A^2). Pep A 1 and Pep A 2 represent the homozygotes and Pep A 2-1 the heterozygote.

Most of the activity in Pep A 1 and Pep A 2 occurs in a single main electrophoretic zone, though this differs in mobility in the two types (fig. 2.3). In the heterozygote however three main zones of activity are seen, two corresponding in mobility to the single zones seen in the homozygotes, whereas the third is exactly intermediate. This rather characteristic type of triple-banded electrophoretic pattern has also been observed in heterozygotes for alleles determining variant forms of a number of other enzymes (Shaw 1965). It is most simply accounted for by the hypothesis that in the homozygous state the enzyme protein contains two identical polypeptide subunits. In the case of peptidase A for example, one may suppose that the allele *Pep* A^1 codes for a polypeptide α^1, and the allele *Pep* A^2 codes for a polypeptide α^2, which since the allelic difference is likely to represent a single mutational step, possibly only differs from α^1 in a single aminoacid substitution. In the heterozygote where both alleles are present, one expects both types of polypeptide (α^1 and α^2) to be formed, and if they combine at random to give dimers, then three distinct enzyme proteins will occur (fig. 2.4). Two of these with subunit structures $\alpha^1\alpha^1$ and $\alpha^2\alpha^2$ will correspond to the forms seen separately in the homozygotes. The third will be a hybrid isozyme, $\alpha^1\alpha^2$, peculiar to the heterozygous state.

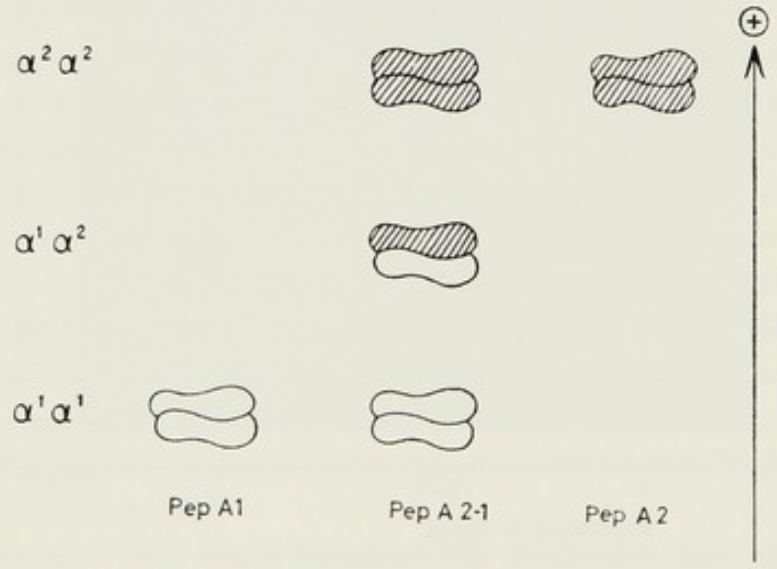


Fig. 2.4. Diagram showing postulated subunit composition of peptidase A types 1, 2-1 and 2.

This interpretation in the case of peptidase A is supported by the results of 'hybridisation' experiments *in vitro*. If solutions of Pep A 1 and Pep A 2 preparations are simply mixed and then subjected to electrophoresis, only the two zones corresponding to those seen in the separate types are observed. However if the mixture is treated under appropriate conditions with urea in the presence of mercaptoethanol, and then after dialysis subjected to electrophoresis, a triple-banded pattern essentially the same as that seen in the heterozygous type Pep A 2-1 is obtained (Lewis and Harris 1969a). The treatment evidently results in the dissociation of the enzyme proteins into their subunits (α^1 and α^2) which subsequently recombine at random to give the hybrid ($\alpha^1\alpha^2$) as well as the original forms ($\alpha^1\alpha^1$ and $\alpha^2\alpha^2$).

The other variant peptidase A types (referred to as Pep A 3-1, Pep A 4-1, Pep A 5-1 etc., fig. 2.5) are very uncommon. Family studies indicate that they are due to a series of rare alleles, each different phenotype representing a heterozygote for one of these rare alleles and for the common allele *Pep* A^1. They show triple-banded electrophoretic patterns, one of the components in each case being the same as the single component in the Pep A 1 phenotype. It seems likely that these different rare alleles each determine a characteristic polypeptide subunit (α^3, α^4, α^5 etc.), and the middle band in each phenotype is a hybrid of this and the α^1-subunit produced by the common allele *Pep* A^1. The other band presumably corresponds to the form of the enzyme which

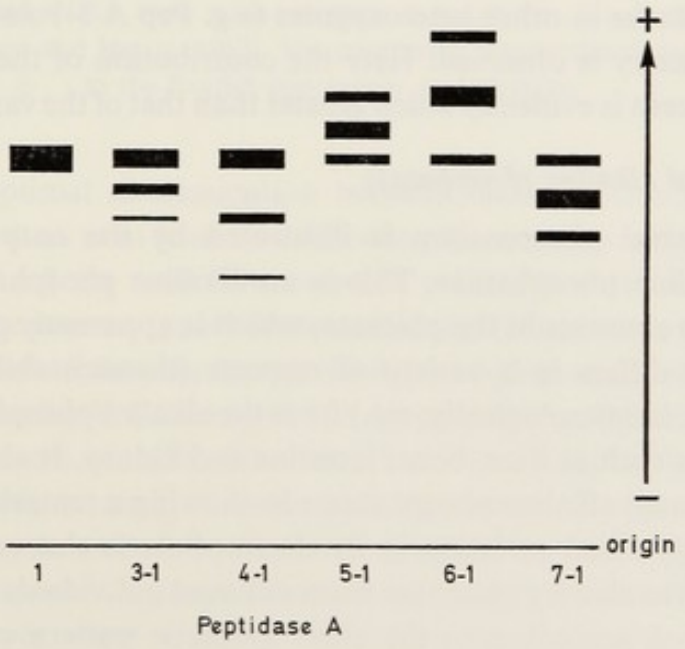


Fig. 2.5. Rare peptidase A types: 3-1, 4-1, 5-1, 6-1 and 7-1. Diagram of main electrophoretic components seen in haemolysates (Lewis and Harris 1967, 1969a, Lewis et al. 1968).

would occur by itself in the homozygote for the particular allele. But since these different alleles are individually relatively infrequent one would expect the corresponding homozygotes to be extremely rare, and indeed none of them have as yet been found.

In heterozygotes where two polypeptide subunits appear to combine in pairs to form three distinct molecular forms of an enzyme, the relative amounts of the three components seen in the electrophoretic pattern often vary considerably from one case to another. This may be because the rates of synthesis or the rates of breakdown of the two polypeptides are not equal, or because the effects on specific activity of the two different polypeptides are not the same. If, for instance, the polypeptide subunits α^x and α^y are formed in equal quantities, have equal activities and combine at random, then the three forms $\alpha^x\alpha^x$, $\alpha^x\alpha^y$ and $\alpha^y\alpha^y$ should occur in the ratio 1:2:1, and the electrophoretic pattern will appear symmetrical, the outer components ($\alpha^x\alpha^x$ and $\alpha^y\alpha^y$) each accounting for 25% of the total activity and the middle or hybrid component ($\alpha^x\alpha^y$) for 50%. However if α^x and α^y differ in their contributions, an asymmetrical electrophoretic pattern will be expected. Thus if the contribution from α^x is twice that from α^y, the three forms $\alpha^x\alpha^x$, $\alpha^x\alpha^y$ and $\alpha^y\alpha^y$ will appear in the ratio 4:4:1. If the contribution of α^x is three times that of α^y then the ratio will be 9:6:1 and so on. Fig. 2.5 illustrates the point with respect to the relative amounts of the three components seen in different peptidase A heterozygotes. In some the pattern appears symmetrical or nearly so. However in other heterozygotes (e.g. Pep A 3-1 and Pep A 4-1) a marked asymmetry is observed. Here the contribution of the α^1 subunit to the overall pattern is evidently much greater than that of the variant (α^3 or α^4).

2.1.3. Placental alkaline phosphatase

The same general phenomenon is illustrated by the enzyme known as placental alkaline phosphatase. This is an alkaline phosphatase found in relatively large amounts in the placenta, which is apparently peculiar to this organ since it differs in a variety of respects (thermostability, inhibition characteristics, immunologically, etc.) from the alkaline phosphatases present in other tissues such as liver, bone, intestine and kidney. It also differs from these other human alkaline phosphatases in showing a remarkable degree of genetically determined variation in its electrophoretic characteristics.

It is possible to classify placentae from different individuals into a number of distinct types according to the electrophoretic pattern of the alkaline phosphatase components or isozymes which they contain (Boyer 1961, Robson and Harris 1965, 1967). Some of the types observed during the course

of examining several thousand different placentae from women in the English population are shown in fig. 2.6. Six of these, known as S_1, F_1, I_1, S_1F_1, S_1I_1 and I_1F_1 are each relatively common and are found respectively in about 42%, 8%, 1%, 33%, 10% and 4% of placentae from this population group. The other so called 'rare' types are individually much less frequent, though together they account for nearly 2% of all placentae.

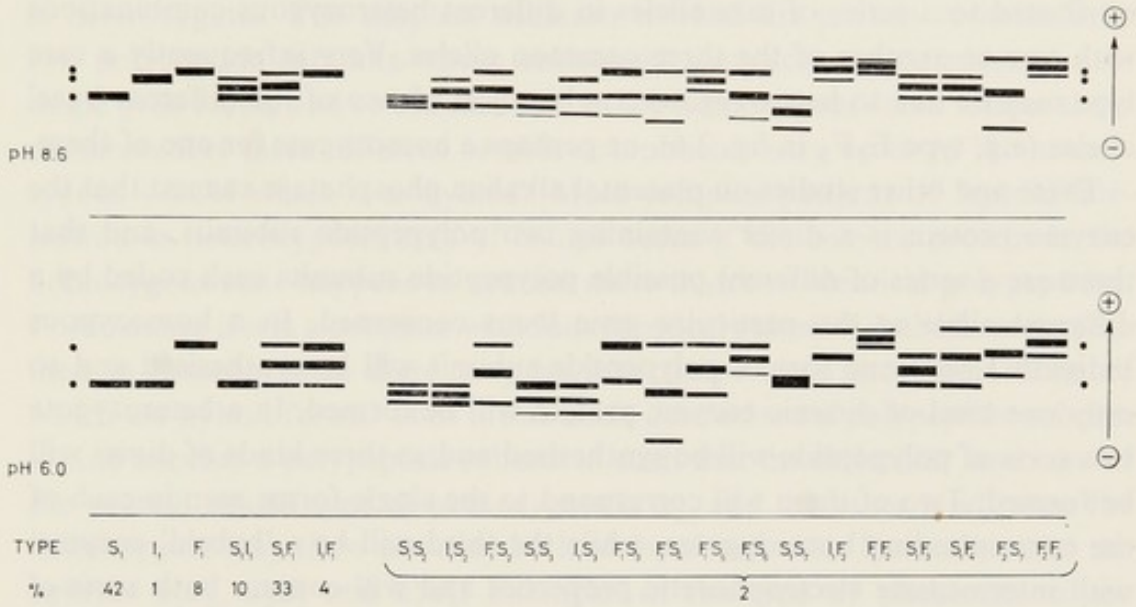


Fig. 2.6. Diagram of electrophoretic patterns of placental alkaline phosphatase of different types, at *p*H 8.6 and *p*H 6.0. Only main components are shown. For further details of the patterns see Robson and Harris (1967). The percentage frequencies of the common types in the English population are indicated.

To obtain optimal discrimination between these various types, electrophoretic separations at at least two different *p*Hs are required, and generally *p*H 8.6 and *p*H 6.0 are used. The essential point about the electrophoretic patterns of six common types illustrated in fig. 2.6, is that while in types S_1, F_1 and I_1 most of the enzyme activity is present in a single electrophoretic component at both *p*H 8.6 and *p*H 6.0, in types F_1S_1, S_1I_1 and F_1I_1 three distinct components are seen either at *p*H 8.6 or *p*H 6.0 or both. The two outer components in each of these triplet patterns correspond in electrophoretic mobility to two of the single components seen in S_1, F_1 or I_1, while the middle component in each case has an exactly intermediate mobility. All the rare types show a triplet pattern of electrophoretic components. Here however while one of the outer components has usually the same mobility as the single component in either S_1, F_1 or I_1 the other is peculiar

to the type. Again the middle band in each case is exactly intermediate in mobility to the outer bands.

Detailed genetical analysis (Robson and Harris 1967) has shown that the six common placental alkaline phosphatase types are determined by three relatively common alleles (Robson and Harris 1965, 1967). Types S_1, F_1 and I_1, represent the three homozygous genotypes, and types F_1S_1, S_1I_1 and F_1I_1, the heterozygous genotypes. The various rare types can be mostly attributed to a series of rare alleles in different heterozygous combinations with one or another of the three common alleles. Very infrequently a rare type may be due to heterozygous combination of two of the different 'rare' alleles (e.g. type F_2F_3 in fig. 2.6), or perhaps a homozygote for one of them.

These and other studies on placental alkaline phosphatase suggest that the enzyme protein is a dimer containing two polypeptide subunits, and that there are a series of different possible polypeptide subunits each coded by a different allele at the particular gene locus concerned. In a homozygous individual only one sort of polypeptide subunit will be synthesised and so only one kind of dimeric enzyme protein will be formed. In a heterozygote two sorts of polypeptide will be synthesised and so three kinds of dimer will be formed. Two of these will correspond to the single forms seen in each of the corresponding homozygotes, while the third will be a 'hybrid' enzyme, with intermediate electrophoretic properties and will contain both sorts of polypeptide.

It is of some interest to compare the relative activities of the three components seen in each of the common heterozygous types F_1S_1, S_1I_1 and F_1I_1. In F_1S_1 the pattern appears symmetrical, the middle component showing roughly twice the activity seen in the two outer components, which are roughly equal. In types S_1I_1 and F_1I_1 however the patterns are asymmetrical, the middle component having a similar activity as one of the outer components, while the other outer component is much weaker. In each case it is the outer component with the I_1 mobility which is weak. This suggests that the polypeptide subunit ($\alpha^I{}_1$) determined by the I_1 allele, contributes less to the observed heterozygous patterns than the polypeptide subunits ($\alpha^S{}_1$ or $\alpha^F{}_1$) determined by the other two common alleles which are approximately equal in this respect. Similar asymmetrical patterns have also been noted in some of the rare heterozygous types. In the case of types S_1I_1 and F_1I_1 there is experimental evidence to suggest that the effect occurs because the $\alpha^I{}_1$ polypeptide subunit is significantly less stable than either $\alpha^S{}_1$ or $\alpha^F{}_1$, and so is broken down more rapidly *in vivo*. Differences in subunit stability may also account for the asymmetrical pattern seen in some of the other rare types,

but it is also possible that kinetic differences or differences in rates of synthesis between the polypeptides coded by the two different alleles present may in some cases account for the effect.

2.1.4. Other enzymes

In general any enzyme or protein which contains multiple polypeptide chains at least two of which are identical in homozygotes may show 'hybrid' forms in heterozygotes. The simplest situation is when the protein contains only two polypeptide subunits whose aminoacid sequences are determined at a single gene locus, and so are identical in the homozygous state. Under these circumstances three different protein components may occur in the heterozygote, one of which is a 'hybrid'. More complex patterns of 'hybrid' formation in heterozygotes may of course occur if the enzyme protein in the homozygous state happens to contain more than two identical polypeptides. For instance if it is a tetramer with four identical subunits in the homozygote then in heterozygotes there may be five components altogether, three of which are hybrid forms. Thus if one allele determined a polypeptide subunit α^x and another a polypeptide subunit α^y such that the molecular structure of the protein in one type of homozygote is α_4^x, and the other is α_4^y, then the heterozygote may show not only both of these molecular forms, but also the hybrids $\alpha_3^x\alpha_1^y$, $\alpha_2^x\alpha_2^y$ and $\alpha_1^x\alpha_3^y$. Examples of this type of phenomenon are provided by lactate dehydrogenase (LDH 1 and LDH 5, see pp. 42–45).

Table 2.1 lists a series of human enzymes classified according to whether in heterozygous individuals the electrophoretic pattern appears to represent simply a mixture of the components separately present in the two corresponding homozygous types, or whether there are in addition one or more 'hybrid' components. In nearly half the cases 'hybrid' enzyme formation appears to occur. This suggests that 'hybrid' formation is probably a relatively common phenomenon, since the enzymes in the table are selected only in that sufficiently precise electrophoretic techniques were available for their study, and that allelic variants happen to have been discovered. Thus as far as their subunit composition and hence their tendency to form hybrids in heterozygotes is concerned, they may be reasonably representative of enzymes in general.

It will be apparent that the demonstration of 'hybrid' components in heterozygotes gives information about the subunit structure of the protein concerned. If such 'hybrids' are demonstrable then the protein must presumably be made up of two or more polypeptide subunits, at least some of which are identical in homozygotes. Enzymes or proteins which contain only a single polypeptide chain will not be expected to show 'hybrid' formation of

TABLE 2.1

Some enzymes classified according to whether or not 'hybrid' isozyme components have been observed in heterozygotes.

(a) Enzymes for which 'hybrid' forms have been observed in heterozygotes	
1. Lactate dehydrogenase	(see pp. 42–45)
2. Placental alkaline phosphatase	(see pp. 30–32)
3. Phosphohexoseisomerase	(Detter et al. 1968)
4. 6-Phosphogluconate dehydrogenase	(Parr 1966)
5. Peptidase A	(see pp. 27–30)
6. Peptidase D (prolidase)	(Lewis and Harris 1969b)
7. Malate dehydrogenase (supernatant enzyme)	(Davidson and Cortner 1967a)
8. Malate dehydrogenase (mitochondrial enzyme)	(Davidson and Cortner 1967b)
9. Red cell 'oxidase' (indophenol oxidase)	(Brewer 1967)
(b) Enzymes for which 'hydrid' forms have *not* been observed in heterozygotes	
1. Phosphoglucomutase	(see pp. 46–53)
2. Adenylate kinase	(Fildes and Harris 1967)
3. Carbonic anhydrase	(Tashian et al. 1968)
4. Adenosine deaminase	(Spencer et al. 1968)
5. Methaemoglobin reductase	(Kaplan and Beutler 1967)
6. Peptidase B	(Lewis and Harris 1967)
7. Red cell acid phosphatase	(see pp. 134–140)
8. Red cell acetylesterase	(Tashian and Shaw 1962)
9. Glucose-6-phosphate dehydrogenase	(see pp. 121–133)

this kind and in such cases the electrophoretic pattern in the heterozygote will represent a simple mixture of the molecular forms seen in the respective homozygotes. However it should be noted that the failure to observe 'hybrids' in heterozygotes cannot be taken by itself as sufficient evidence that the enzyme is a monomer. The absence or apparent absence of hybrids in heterozygotes even though the enzyme is polymeric could be due to several different causes. The enzyme protein may be so readily dissociated into its subunits as occurs for example with haemoglobin, that in the course of the electrophoretic or chromatographic separation of the mixture in heterozygotes the hybrid forms are lost. Alternatively there may be some special structural feature of the alternative polypeptide chains which prevents or limits their association in the same multimeric molecule. A further possible cause is that the polypeptide products of the two alleles although they are both synthesised in the organism may only be formed in different cells so

they do not have the opportunity to combine together in a single protein molecule. This has been shown to be the case for the enzyme glucose-6-phosphate dehydrogenase which is determined by a gene locus on the X chromosome. Here hybrid formation has been demonstrated *in vitro* (Yoshida et al. 1967b), but is not seen in female heterozygotes. This is thought to occur because one or other of the two alleles present in any single cell of the female is not functional, being present on the so-called 'inactivated' X chromosome (pp. 131–134). One may anticipate that the same will be generally true for other multimeric enzymes whose polypeptide chains are coded by gene loci on the X chromosome.

2.2. *The several gene loci determining haemoglobin polypeptide chains*

So far we have been mainly concerned with the effects on protein structure of mutations at single gene loci. We now consider the complexities that arise when two or more separate gene loci are involved in defining the structure of a protein or a closely related set of proteins. Haemoglobin, again, provides a good illustration.

In the normal adult the principal type of haemoglobin present is Hb A, which has already been discussed. In the red cells of the foetus, however, another form predominates. This is known as foetal haemoglobin, or Hb F. Although it differs in its detailed structure from Hb A it resembles it very closely in its three-dimensional conformation and in most of its properties. In the newborn about 70–80% of the haemoglobin present is Hb F, and nearly all the remainder Hb A. This proportion rapidly changes during the first few months after birth, so that by six or twelve months Hb F is present in no more than trace amounts and Hb A now predominates.

Yet a further distinctive molecular form of haemoglobin known as Hb A_2 is also consistently found in normal individuals. It occurs along with Hb A in the red cells of the adult, but it constitutes only a small fraction (about 2.5%) of the total haemoglobin present. It is apparently synthesised synchronously with Hb A but at a much slower rate.

Hb A_2 and Hb F, like Hb A, both contain two distinct types of polypeptide, each of which is represented twice in the molecule. One of these is an α-chain identical with that in Hb A. They differ from Hb A and also from each other in the nature of the other chain, which in Hb A_2 is called the δ-chain and in Hb F the γ-chain. So these three different forms of normal haemoglobin may be written as follows:

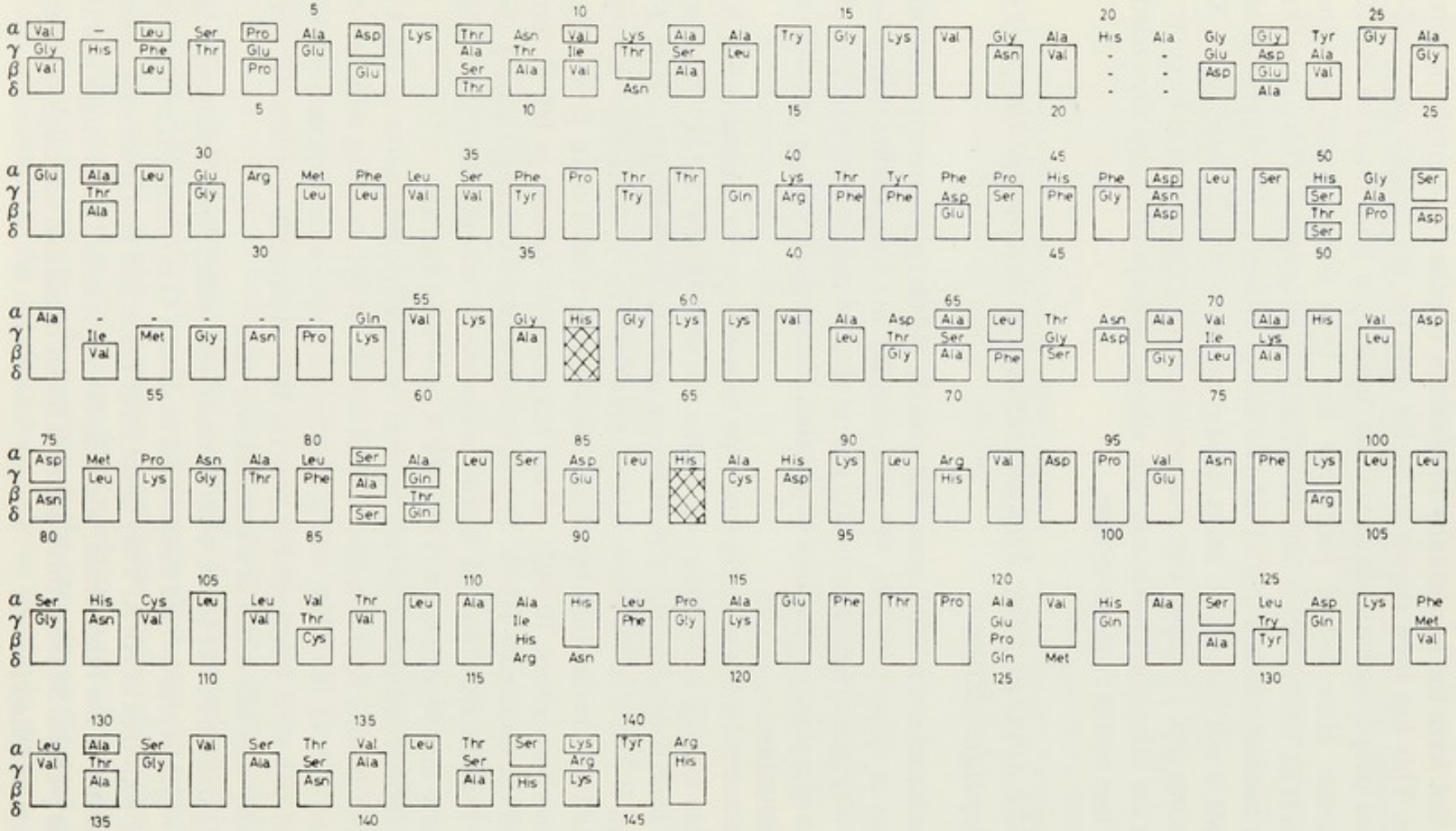


Fig. 2.7. The aminoacid sequences of the α, β, γ and δ polypeptide chains of haemoglobin. Identical aminoacids are boxed. The haem linked histidine residues (p. 20) are hatched. The 141 aminoacids in the α-chain are numbered at the top, and the 146 aminoacids in the β-, γ- and δ-chains are numbered at the bottom. The 'gaps' are arranged to secure maximum homology in the aminoacid sequences. (From Lehmann and Carrell 1969.) Aminoacid abbreviations as in fig. 1.5, p. 13.

Hb A $\alpha_2\beta_2$
Hb A_2 $\alpha_2\delta_2$
Hb F $\alpha_2\gamma_2$

As in the β-chain there are 146 aminoacid residues in the δ- and the γ-chains, and the sequences of the three chains show many similarities (fig. 2.7). The β-chain and the δ-chain sequences have identical aminoacid residues in 136 positions and differ in only ten positions. The γ-chain shows more differences but still shares identical residues in 107 positions with the β-chain, and in 105 positions with the δ-chain. There are also many similarities between the sequences of the β-, δ- and γ-chains on the one hand and the α-chain on the other. Exact comparisons are complicated because the latter chain has five fewer aminoacid residues, but making reasonable assumptions about the positions of the 'gaps', it has been estimated that there is a correspondence of perhaps 46% in the aminoacid sequences of the α- and the β-chains. The remarkable degree of homology in the aminoacid sequences of these different polypeptide chains suggests that they have all been derived during the course of evolution from a single ancestral form. This point is considered in more detail later (ch. 3, pp. 77–79).

Since the α-, β-, γ- and δ-chains all occur in normal individuals, and since they differ from each other in varying degrees in their aminoacid sequences, one must suppose that they are determined at separate gene loci on the chromosomes. Thus at least four gene loci will be concerned in determining the structures of the three normal haemoglobins A, A_2 and F, and the situation may be represented as shown in fig. 2.8. An important conclusion follows.

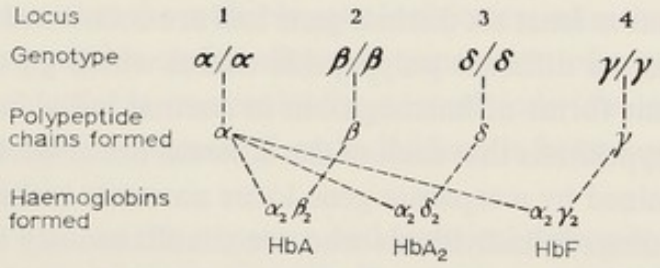


Fig. 2.8. Formation of haemoglobins A, A_2 and F in the normal individual.

A mutant allele at the locus determining the α-chain will be expected to result in the occurrence of variant forms of all three haemoglobins A, A_2 and F. On the other hand a mutant allele at one or other of the loci determining the β-, δ- or γ-chains should only result in a variant form of the haemoglobin

which normally contains the corresponding chain. A number of variants of Hb A_2 and Hb F have indeed been identified, and the findings are fully consistent with these expectations. In individuals heterozygous for a mutant allele at either the β-, δ- or γ-locus, only a variant form of the haemoglobin normally containing that particular polypeptide chain is produced. Thus in the sickle-cell trait where the β-locus is involved, haemoglobin A_2 ($\alpha_2\delta_2$) and haemoglobin F ($\alpha_2\gamma_2$) are normal. On the other hand, an individual heterozygous for a mutant allele at the α-locus synthesises variant forms as well as the normal forms of each of the three kinds of haemoglobin, Hb A ($\alpha_2\beta_2$), Hb A_2 ($\alpha_2\delta_2$) and Hb F ($\alpha_2\gamma_2$).

In early embryonic life it seems that still another form of haemoglobin polypeptide chain is synthesised (Huehns et al. 1964). This has been called the ε-chain and although its sequence has not yet been determined in detail, it evidently differs from those of the α-, β-, γ- or δ-chains. Several so-called 'embryonic' haemoglobins with structures thought to be ε_4, $\alpha_2\varepsilon_2$ and $\gamma_2\varepsilon_2$ have been found in significant quantities in early embryos and have also been detected in very small amounts in later foetal life. Such haemoglobins are of course peculiarly difficult to study and so far no variants have been detected. However, it seems very probable that this chain is determined by yet a further gene locus distinct from the loci which code for the α-, β-, γ- and δ-chains.

Furthermore it has also been shown that two slightly different sorts of γ-chain are formed by the normal organism. In one of these, position 136 in the aminoacid sequence is occupied by glycine (as shown in fig. 2.8), while in the other this position is occupied by alanine (Schroeder et al. 1968). The evidence suggests that separate gene loci code for these two different γ-chains. If so it seems that at least six distinct gene loci are concerned in determining the structures of the different polypeptide chains which go to make up the various molecular forms of haemoglobin in normal individuals.

The general hypothesis that each of the different haemoglobin polypeptide chains is determined by a separate gene locus accounts rather nicely for the biochemical findings in individuals who are simultaneously heterozygous at the locus which determines the α-chain and also at one of the other loci. Under such circumstances a particularly complex mixture of haemoglobins is produced (Baglioni and Ingram 1961, McCurdy et al. 1961).

As an example (Weatherall et al. 1962) one may consider an individual who is heterozygous at the α-locus for the mutant allele which determines the variant polypeptide chain α^G ($\alpha^{68\ \text{Asn} \rightarrow \text{Lys}}$), and is also heterozygous at the β-locus for the Hb C gene which determines the variant polypeptide chain

β^C ($\beta^{6\ Glu \rightarrow Lys}$). Two sorts of α-chain are synthesised and two sorts of β-chain, and as a result four different kinds of haemoglobins containing α- and β-type chains are demonstrable by electrophoresis (fig. 2.9). These are

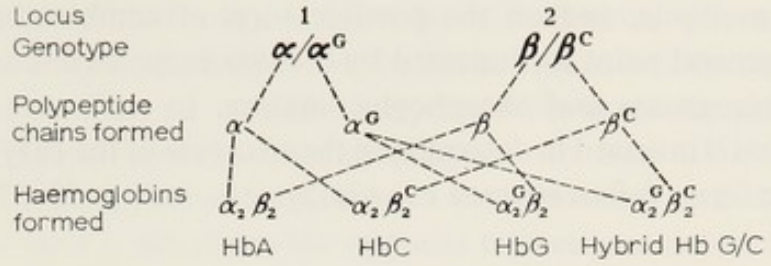


Fig. 2.9. Formation of haemoglobins A, C, G and G/C in an individual heterozygous at the α locus for an allele which determines the variant polypeptide α^G, and heterozygous at the β locus for an allele which determines the polypeptide β^C.

normal Hb A ($\alpha_2\beta_2$), Hb C ($\alpha_2\beta_2^C$), Hb G ($\alpha_2^G\beta_2$), and a new haemoglobin, $\alpha_2^G\beta_2^C$, which contains both the variant α-chain and also the variant β-chain.

Also, of course, in such an individual a small fraction of the haemoglobin present in adult life will consist of Hb A_2 ($\alpha_2\delta_2$) and the variant Hb A_2^G ($\alpha_2^G\delta_2$); while in foetal life and immediately after birth most of the haemoglobin formed will consist of Hb F ($\alpha_2\gamma_2$), and the variant Hb F^G ($\alpha_2^G\gamma_2$), and since there are two sorts of γ-chain there will presumably be two sorts of each of these kinds of molecule. Besides this, various 'hybrid' forms of haemoglobin molecules, such as $\alpha_2\beta\beta^C$ and $\alpha\alpha^G\beta_2$, will also be present in the red cells of such individuals, though they will not be demonstrable in ordinary electrophoretic or chromatographic separation of haemolysates for the reasons already discussed.

Thus one can see that the multiple gene loci determining the several distinct haemoglobin polypeptide chains allow the generation of a quite complex set of different haemoglobin molecules in normal individuals, and that this complexity is greatly enhanced if an individual is heterozygous at one or more of these loci.

More is known about the structure and genetical determination of haemoglobin than any other human protein. But for a number of other proteins there is now good, though much less complete evidence that two or more gene loci are concerned in determining their structures. The phenomenon, though common, is not universal, since there are also many proteins which appear to contain only one type of polypeptide chain apparently defined by just a single gene locus.

Where an enzyme or protein is determined by two or more gene loci each

coding for the aminoacid sequence of a distinctive polypeptide chain, several molecular forms of the protein will often be demonstrable in the normal individual. Their number, properties and characteristics will depend on whether or not the different polypeptide products of the several loci combine in a single molecule, and on the possible sorts of combination that can occur. The general point is illustrated by the two enzymes discussed below: lactate dehydrogenase *and* phosphoglucomutase. In each case more than one gene locus is involved in determining the structure of the enzyme protein, and multiple forms (often referred to as isozymes – see pp. 53–57) occur.

2.3. Lactate dehydrogenase

Lactate dehydrogenase (L-lactate : NAD oxidoreductase) is a key enzyme in carbohydrate metabolism and occurs in virtually all tissues. It catalyses the interconversion of lactate and pyruvate with the concomitant oxidation and reduction of the coenzyme NAD:

$$\underset{\text{pyruvic acid}}{CH_3, CO.COOH} + NADH_2 \rightleftarrows \underset{\text{lactic acid}}{CH_3, CHOH.COOH} + NAD$$

When extracts from various tissues are examined by electrophoresis, at least five distinct proteins (isozymes) with lactate dehydrogenase activity can be separated. The relative amount of the several isozymes is found to vary

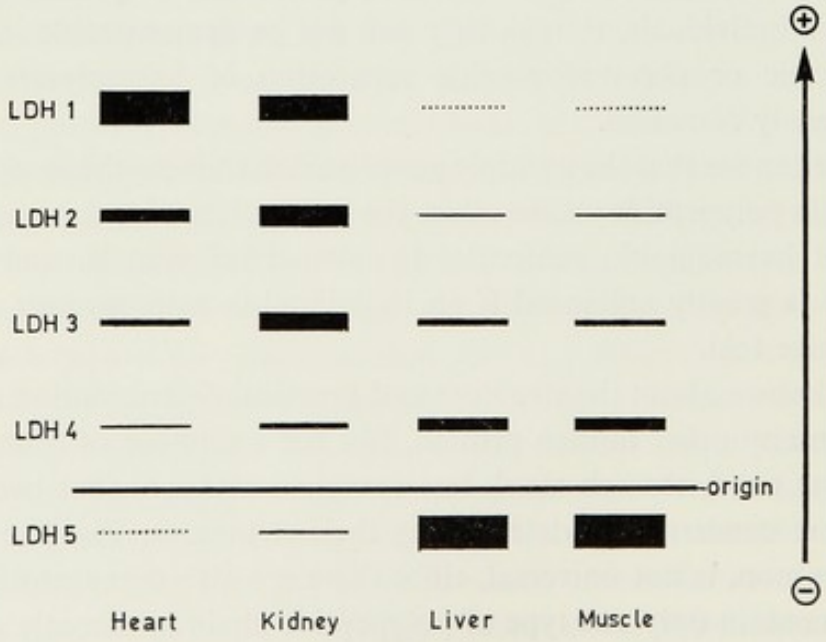


Fig. 2.10. Lactate dehydrogenase isozymes in heart muscle, kidney, liver, and skeletal muscle.

markedly from tissue to tissue. This is illustrated in fig. 2.10 where it can be seen, for instance, that the electrophoretic pattern of lactate dehydrogenase observed in heart muscle and in kidney is strikingly different from that observed in liver or in skeletal muscle. The five isozymes are referred to as LDH 1, LDH 2, LDH 3, etc. in the order of their electrophoretic mobilities towards the anode at alkaline *p*H. In heart muscle and in kidney LDH 1 predominates. The other isozymes are present but in much smaller amounts, and their relative activities decrease progressively from LDH 2 to LDH 5. In skeletal muscle and in liver, LDH 5 predominates, and the pattern of relative activities is effectively the opposite to that seen in heart.

It has been shown that each of these lactate dehydrogenase proteins are tetramers composed of four polypeptide subunits (Appella and Markert 1961, Markert 1963, 1968), and the polypeptide subunits can be of two different kinds which are usually referred to as A and B (or M and H). Although the detailed structures are not yet known, it appears from comparisons of their aminoacid compositions and their peptide fingerprints that they probably differ quite extensively in aminoacid sequence, though they are evidently similar in molecular size. Extensive evidence has been obtained to show that the five characteristic isozymes have subunit structures which may be represented as follows: LDH 1 = B_4, LDH 2 = AB_3, LDH 3 = A_2B_2, LDH 4 = A_3B and LDH 5 = A_4.

The LDH isozyme patterns characteristic of different tissues can be largely accounted for, if it is supposed that although both the A and B subunits are usually formed they occur in widely different amounts in different tissues, and the relative quantities of the several isozymes which are observed are the consequence of random combinations of the subunits present. Thus in skeletal muscle subunit A is evidently in great excess compared with subunit B, so that LDH 5 is the predominant form and the others are present in decreasing amounts in the order LDH 4 > LDH 3 > LDH 2 > LDH 1. Heart muscle by contrast has a great excess of the B subunit, so that here LDH 1 > LDH 2 > LDH 3 etc. It is possible experimentally (Markert 1963a) to dissociate the separate isozyme proteins into their constituent subunits and then allow them to recombine to give enzymically active products (fig. 2.11). Applying this to artificial mixtures of purified LDH 1 (B_4) and LDH 5 (A_4) in different proportions, isozyme patterns corresponding to those observed in different tissues can be generated. From such experiments (Vessell 1965a) it appears that the ratio of A to B subunits in human skeletal muscle and liver cells may be of the order 10:1. For heart muscle on the other hand the ratio appears to be close to 1:20, and for kidney 1:10.

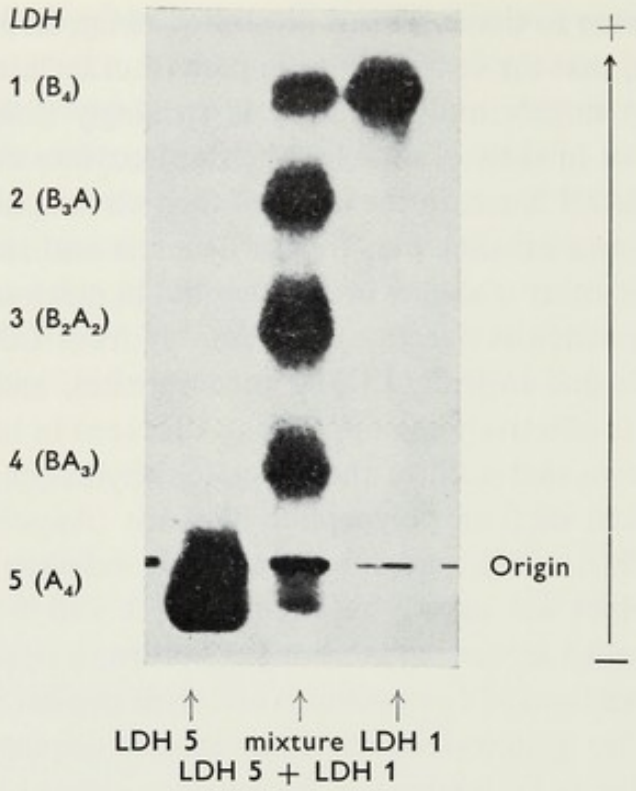


Fig. 2.11. Experiment showing the dissociation and recombination of lactate dehydrogenase subunits (Markert 1963a). Preparations of LDH 1 (B_4) and LDH 5 (A_4), separately and mixed together in equal proportions, were placed in 1 M NaCl and frozen overnight. The electrophoretic patterns obtained after thawing the samples are shown. In the mixed preparation the five isozymes LDH 1 (B_4), LDH 2 (B_3A), LDH 3 (B_2A_2), LDH 4 (BA_3) and LDH 5 (A_4) were present in the approximate ratio 1:4:6:4:1.

Numerous investigations have been directed to comparing the properties of the various isozymes. Significant differences between LDH 1 (B_4) and LDH 5 (A_4) have been demonstrated in K_m's, in reactivity with various analogues of the coenzyme NAD, in inhibition by substrate excess, in thermostability and in other respects. These differences are presumably reflected functionally, and one would expect the pattern of isozymes present in a given tissue to be related to its particular metabolic characteristics. However, although certain general correlations have been noted, for example the preponderance of A subunits in tissues which tend to be subjected to periods of oxygen lack (Cahn et al. 1962, Wilson et al. 1963), the full details of such possible functional relationships remain to be elucidated (Vessell 1968, Kaplan et al. 1968).

One may anticipate that the two distinct polypeptide subunits A and B are determined by separate gene loci. If so, a mutation would be expected to

produce quite different effects on the LDH isozymes according to whether it occurred at the locus determining the A polypeptide, or at the locus determining the B polypeptide. Thus a mutant allele at the '*A*' locus should not affect the LDH 1 protein because this does not contain an A subunit. It should however affect in different degrees LDH 2, LDH 3, LDH 4 and LDH 5, which contain respectively one, two, three and four A type subunits. Similarly a mutant allele at the '*B*' locus should not affect LDH 5 but should affect the four other isozymes.

It is apparent that very complex sets of isozymes may be generated in heterozygotes, and it is possible to predict in some detail the general pattern of variant proteins to be expected in different cases. Thus an individual heterozygous at the '*A*' locus but homozygous for the normal allele at the '*B*' locus would be expected to form the fifteen isozymes shown in column (c) of table 2.2, while an individual heterozygous at the '*B*' locus but homozygous at the '*A*' locus would be expected to form the isozymes shown in column (d) of the table. The relative proportions of the normal and variant isozymes formed will be determined by the amounts of the normal and variant polypeptides available to form tetramers. If for example in the case of an '*A*' locus mutation, equal amounts of the variant 'A' polypeptide and the normal A polypeptide were formed, and if association to give tetramers occurred at random then the two isozymes corresponding to LDH 2 should occur in the ratio 1:1, the three isozymes corresponding to LDH 3 in the ratio 1:2:1, the four isozymes corresponding to LDH 4 in the ratio 1:3:3:1, and the five isozymes corresponding to LDH 5 in the ratio 1:4:6:4:1. These ratios would be consistent from tissue to tissue, but the overall pattern observed in any given tissue would of course be correlated with the relative amounts of the five isozymes normally present, that is the relative activities attributable to the two loci ('*A*' and '*B*').

A number of different genetically determined variants of lactate dehydrogenase have indeed been discovered (Boyer et al. 1963, Kraus and Neely 1964, Nance et al. 1963, Davidson et al. 1965, Vessell 1965b.) They are each relatively uncommon and they have mainly been found during the course of routine electrophoretic surveys in various populations. None of these so far discovered appears to be specifically associated with any clinical abnormality. In each of these variants a complex but quite characteristic electrophoretic pattern of many isozymes is seen. This runs true to type in the particular family in which it is found and the familial distribution indicates that the individuals showing the peculiarity must be heterozygous at an autosomal locus for a rare mutant gene and its normal allele. Because of the rarity

of these mutants in the populations that have been studied, no homozygotes have as yet been identified.

From the electrophoretic patterns it is possible to classify these different

TABLE 2.2

Postulated subunit constitutions of LDH isozymes in normal individuals and in individuals heterozygous for mutant genes at either the '*A*' or '*B*' loci.

(a) Isozyme	(b) Normal subunit constitution	(c) Subunit constitution of the fifteen possible isozymes in a heterozygote for an '*A*' locus variant	(d) Subunit constitution of the fifteen possible isozymes in a heterozygote for a '*B*' locus variant
LDH 1	B_4	B_4	B_4 B_3B^* $B_2B^*_2$ BB^*_3 B^*_4
LDH 2	B_3A	B_3A B_3A^*	B_3A B_2B^*A BB^*_2A B^*_3A
LDH 3	B_2A_2	B_2A_2 B_2AA^* $B_2A^*_2$	B_2A_2 BB^*A_2 $B^*_2A_2$
LDH 4	BA_3	BA_3 BA_2A^* BAA^*_2 BA^*_3	BA_3 B^*A_3
LDH 5	A_4	A_4 A_3A^* $A_2A^*_2$ AA^*_3 A^*_4	A_4

A and B represent the normal subunits.
A* and B* represent the variant subunits.

heterozygous types quite readily into distinct classes, according to whether the rare mutant allele is at the '*A*' locus or at the '*B*' locus, and so determines a variant form of the A polypeptide or the B polypeptide. The point is illustrated in fig. 2.12, which shows the electrophoretic patterns observed in

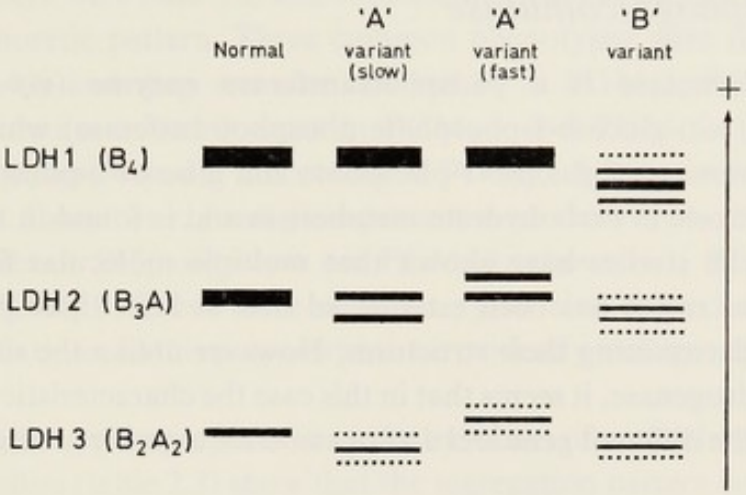


Fig. 2.12. Diagram of variant electrophoretic patterns of lactate dehydrogenase as seen in red cells. Two of the variants are from individuals heterozygous for different alleles at the gene locus determining the A subunit, and the other is from an individual heterozygous at the gene locus determining the B subunit.

red cell lysates from different types of heterozygotes. In normal red cells the three main isozymes found are LDH 1, LDH 2 and LDH 3. It will be seen that the patterns observed in one of the variant types corresponds exactly to that expected in a heterozygote for a '*B*' locus mutant, and that the other two agree with the expectations for '*A*' locus mutants. Similar mutants have also been observed in other species (e.g. *Peromyscus*, Shaw and Barto 1963). Here the homozygote for the mutant could also be studied and the findings in different tissues were as expected on the hypothesis.

Besides the gene loci giving rise to the characteristic A and B subunits of LDH, there is evidence for yet a third locus determining a distinctive subunit (Bianco and Zinkham 1963, Zinkham 1968). This so-called C subunit leads to the formation of a further type of tetrameric isozyme (usually referred to as the X band), which appears to be peculiar to sperm cells. The isozyme has been observed in man and in various other species, but inherited variants have so far only been demonstrated in pigeons (Bianco et al. 1964). Apparently the C subunit is formed only in the primary spermatocyte and then only for a short time. Although isozymic tetramers containing the C subunit as well as

the A or B subunits may be generated *in vitro* by appropriate dissociation and recombination experiments, they do not occur *in vivo*. This is presumably because the A or B subunits do not occur together with the C subunit when it is being synthesised in the primary spermatocyte.

2.4. Phosphoglucomutase

Phosphoglucomutase is a phosphotransferase enzyme (α-D-glucose-1,6-diphosphate; α-D-glucose-1-phosphate phosphotransferase) which catalyses the interconversion of glucose-1-phosphate and glucose-6-phosphate. It has an important role in carbohydrate metabolism and is found in most tissues. Electrophoretic studies have shown that multiple molecular forms of the enzyme occur and it has been established that at least three gene loci are involved in determining their structures. However unlike the situation with lactate dehydrogenase, it seems that in this case the characteristic polypeptide products of the different gene loci do not associate together in single isozymes.

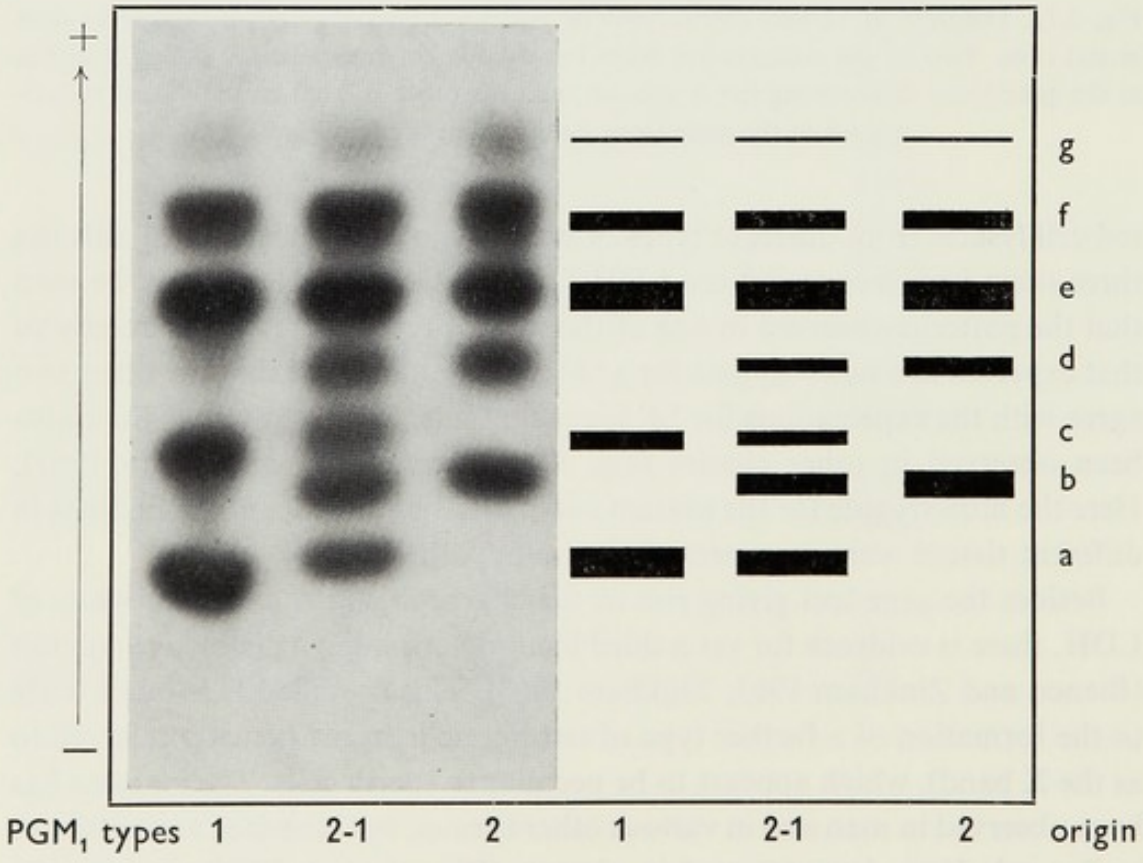


Fig. 2.13. Photograph and diagram of red cell phosphoglucomutase isozymes in the three common PGM_1 types: PGM_1 1, PGM_1 2-1 and PGM_1 2. Electrophoresis in starch gel at *p*H 7.4 (Spencer et al. 1964a).

The complexity of the phosphoglucomutase isozymes first became apparent in studies in human red cells obtained from different individuals (Spencer et al. 1964a). When red cell lysates were subjected to starch gel electrophoresis and a staining procedure which specifically detected phosphoglucomutase activity was applied, multiple zones of activity were demonstrable, and it was found that there were clear cut and consistent person to person differences in the electrophoretic pattern. Three common phenotypes were readily recognised (fig. 2.13) and they are now referred to as $PGM_1 1$, PGM_1 2–1 and PGM_1 2. Isozymes *a* and *c* occur in PGM_1 1 and PGM_1 2–1, but not in PGM_1 2. Isozymes *b* and *d* are absent in PGM_1 1 but occur in PGM_1 2–1 and PGM_1 2. Isozymes *e*, *f* and *g* are present in all three types. In the English population about 58% of people are PGM_1 1, about 36% PGM_1 2–1 and about 6% PGM_1 2. These three common types also occur in many other human populations, so this polymorphism is evidently widespread throughout the species.

Studies of these phenotypes among the members of a very large number of different families (table 2.3) show that the segregation pattern conforms very

TABLE 2.3

The distribution of phosphoglucomutase types (PGM_1) in 537 families (Harris et al. 1968).

Parents	Number of matings	Children			Total
		1	2–1	2	
1 × 1	199	392	—	—	392
1 × 2–1	203	207	215	—	422
1 × 2	34	—	71	—	71
2–1 × 2–1	77	35	81	41	157
2–1 × 2	21	—	24	31	55
2 × 2	3	—	—	13	13
Totals	537	634	391	85	1110

closely to Mendelian expectations on the simple hypothesis that the polymorphism is determined by two common autosomal alleles (PGM_1^1 and PGM_1^2). Phenotypes PGM_1 1 and PGM_1 2 represent the homozygous genotypes $PGM_1^1PGM_1^1$ and $PGM_1^2PGM_1^2$, and phenotype PGM_1 2–1 the heterozygous

genotype $PGM_1^1PGM_1^2$. Thus the isozymes *a* and *c* appear to be determined by one allele (PGM_1^1), and the isozymes *b* and *d* by the other (PGM_1^2). In the heterozygote where both alleles are present all four isozymes occur, but there is no evidence in this case of 'hybrid' isozyme formation.

Following the discovery of these common phosphoglucomutase types, blood samples from several thousand different individuals were examined, and in the course of this work a number of other isozyme patterns each characteristic of the individual, but each very rare, were discovered (Hopkinson and Harris 1966). Pedigree studies show that many of these can be attributed to the heterozygous combination of either PGM_1^1 or PGM_1^2 with one or another of a series of rare alleles at the same gene locus. Some examples of such isozyme patterns are shown in fig. 2.14. In each case there are two

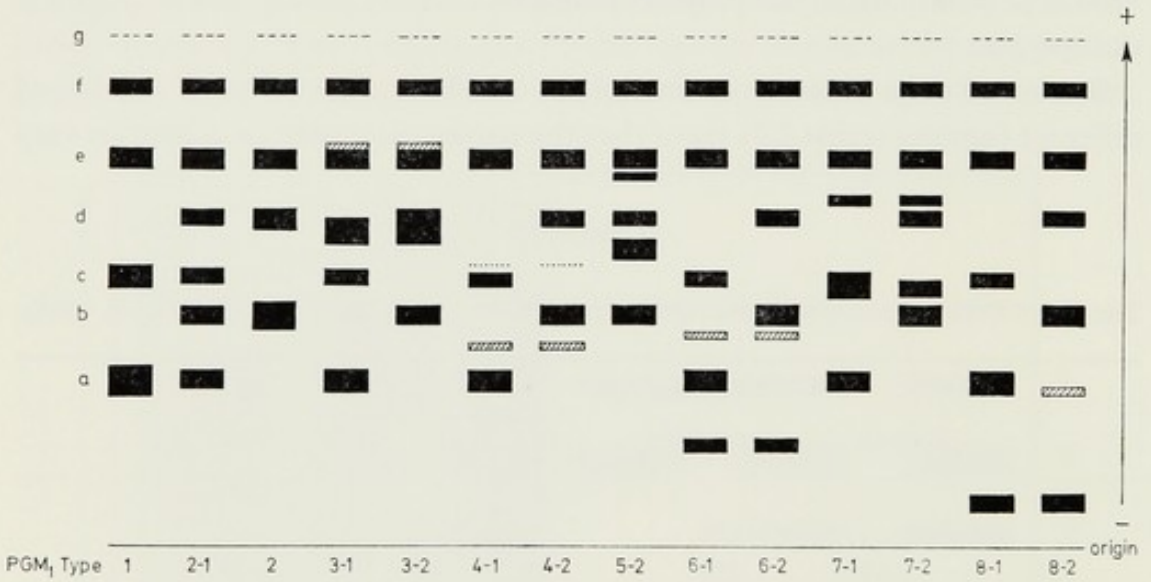


Fig. 2.14. Diagram showing red cell phosphoglucomutase isozymes observed in 14 different PGM_1 types (Hopkinson and Harris 1966).

unusual isozymes presumably determined by the particular rare allele, and also either the *a* and *c* isozymes determined by PGM_1^1, or the *b* and *d* isozymes determined by PGM_1^2.

The fact that isozymes *e*, *f* and *g* are not apparently affected by the common or rare alleles mentioned so far, suggests that they are determined by a second gene locus (PGM_2). Direct evidence for this (Hopkinson and Harris 1965, 1966) came from the recognition of several rare variant types which appeared to involve *e*, *f* and *g* and to occur independently of the variants attributable to the first locus (PGM_1). Several examples of these PGM_2

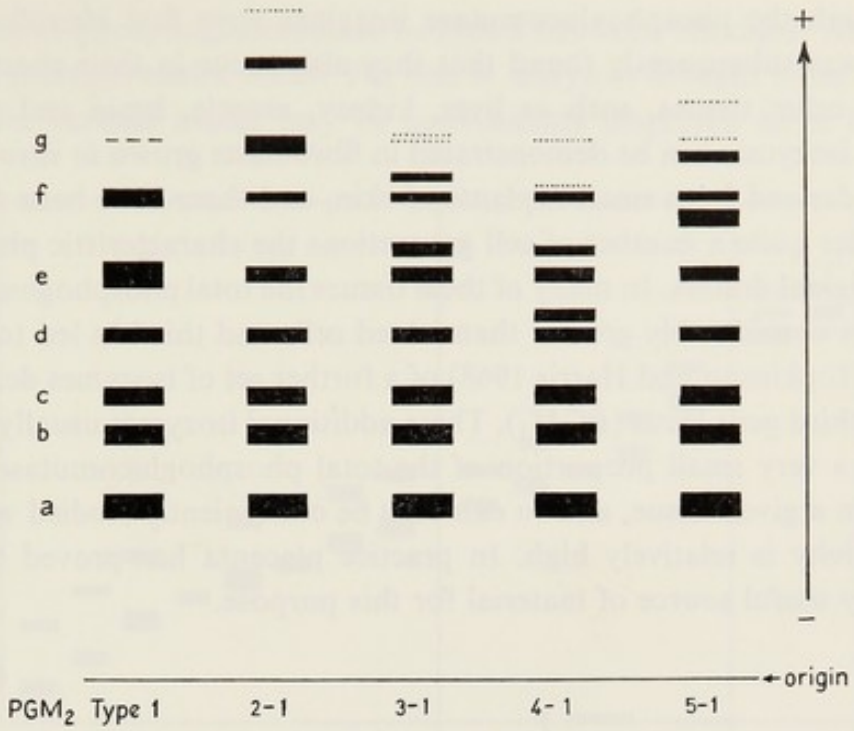


Fig. 2.15. Diagram showing red cell phosphoglucomutase isozyme patterns in five different PGM_2 types. In each case the PGM_1 type shown is PGM_1 2-1 (Hopkinson and Harris 1966; Parrington et al. 1968.)

variant isozyme patterns are illustrated in fig. 2.15. They are shown together with the *a*, *b*, *c* and *d* isozymes characteristic of PGM_1 2–1, but they may also be associated with *a* and *c* alone (i.e. with PGM_1 1) or with *b* and *d* alone (i.e. with PGM_1 2).

From pedigree studies one may infer that these different variant *e*, *f*, *g* patterns occur in individuals heterozygous for one or another of several uncommon alleles at a second gene locus (PGM_2). The usual allele at this locus (PGM_2^1) for which most people are homozygous, is thought to determine isozymes *e*, *f* and *g*, and the different mutant alleles to result in altered forms of these isozymes with characteristic and consistent changes of electrophoretic mobility. The patterns seen in the different heterozygotes can be interpreted as due to a simple mixture of the isozymes determined by the two alleles present. In some cases (e.g. PGM_2 3–1) the new isozymes determined by the mutant allele are clearly separated from isozymes *e*, *f* and *g* determined by the usual allele (PGM_2^1). But in other cases the two groups may overlap to some extent. For example in PGM_2 2–1, the slowest isozyme of the variant group appears to have about the same electrophoretic mobility as *g*, the fastest of the three isozymes determined by PGM_2^1.

Although the phosphoglucomutase isozymes were first identified in red cells, it was subsequently found that they also occur in their characteristic types in other tissues, such as liver, kidney, muscle, brain and placenta. Also the isozymes can be demonstrated in fibroblasts grown *in vitro* in tissue cultures derived from small explants of skin, and these have been shown to retain after quite a number of cell generations the characteristic phenotypes of the original donors. In many of these tissues the total phosphoglucomutase activity is considerably greater than in red cells and this has led to the discovery (Hopkinson and Harris 1968) of a further set of isozymes determined by yet a third gene locus (PGM_3). These additional isozymes usually account for only a very small proportion of the total phosphoglucomutase activity present in a given tissue, and so can only be conveniently studied where the total activity is relatively high. In practice placenta has proved to be an extremely useful source of material for this purpose.

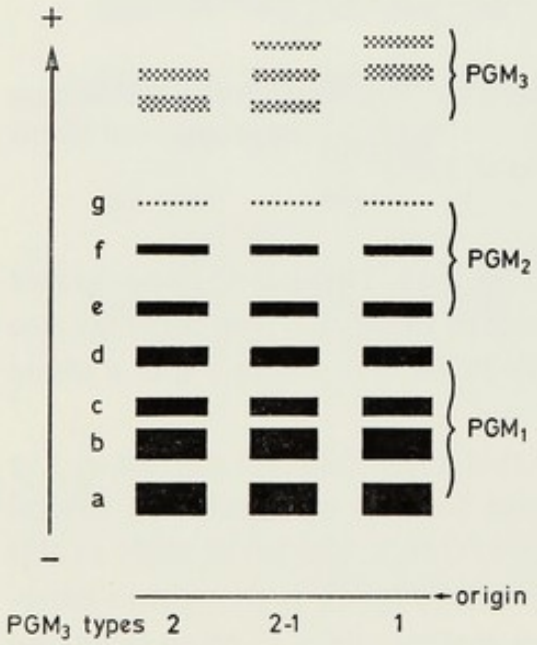


Fig. 2.16. Diagram of three types of PGM_3 isozymes, PGM_3 1, PGM_3 2-1 and PGM_3 2, as seen in placental extracts. In each case the PGM_1 isozymes are PGM_1 2-1, and the PGM_2 isozymes are PGM_2 1 (Hopkinson and Harris 1968).

This third set of isozymes migrate more rapidly towards the anode than the others, and the typical electrophoretic patterns that are observed are shown in fig. 2.16. Three distinct phenotypes have been recognised (PGM_3 1, PGM_3 2–1 and PGM_3 2) and genetical analysis indicates that they are determined by two common alleles PGM_3^1 and PGM_3^2, the phenotype PGM_3 2–1 representing the heterozygote. The types occur independently of those determined by the alleles at the other loci.

Thus evidence for three different gene loci separately determining a

distinct set of phosphoglucomutase isozymes has been obtained. At each of these loci multiple alleles occur. The sets of isozymes thought to be attributable to the various alleles may be conveniently displayed as in fig. 2.17.

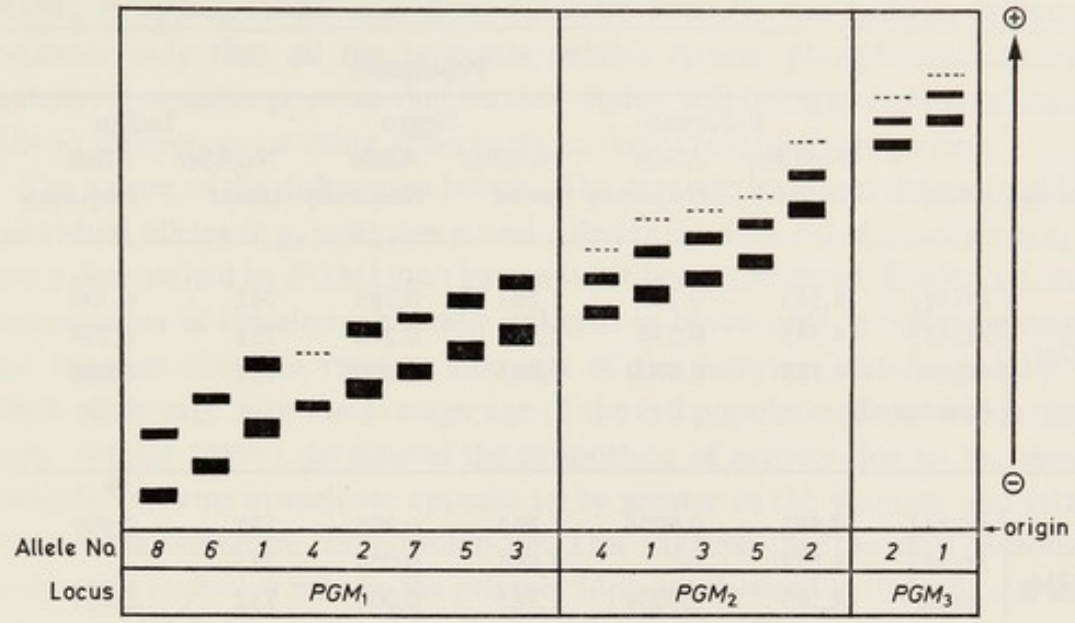


Fig. 2.17 Diagram of phosphoglucomutase isozymes determined by eight different alleles at locus PGM_1, five alleles at locus PGM_2, and two alleles at locus PGM_1 (Hopkinson and Harris 1969). At least one additional, relatively weak, fast moving isozyme probably also occurs consistently in the PGM_1 series, but is generally obscured by the other isozymes present in the different genotypes.

Individuals homozygous at all three loci show at least eight phosphoglucomutase isozymes. In individuals heterozygous at one or more loci the isozyme patterns are more complex, and it is apparent that a very large number of different combinations of phosphoglucomutase isozymes are possible in different individuals. Their relative incidence will of course depend on the frequencies of the different alleles in the particular population (table 2.4), and most of the possible allelic combinations are very rare. Nevertheless one can classify people in the English population for example, into at least eight distinct types, each of which has an appreciable frequency (greater than 2%) in the general population (table 2.5). Thus quite a high degree of individual differentiation occurs.

So far not much is known about the nature of the differences between the isozyme products of the three separate loci. Studies by gel filtration (Monn 1969a) and ultracentrifugation (Santachiara and Modiano 1969) suggest that the PGM_2 isozymes have a somewhat larger molecular size than the

TABLE 2.4

Estimated frequencies of phosphoglucomutase alleles in population samples from different ethnic groups. (Data collected by MRC Human Biochemical Genetics Unit, Galton Laboratory, London.)

		Population					
		European		Negro		Indian	
Locus	Allele	Number tested	Allele frequency	Number tested	Allele frequency	Number tested	Allele frequency
PGM_1	PGM_1^1	4,583	0.765	1,863	0.785	732	0.700
	PGM_1^2	4,583	0.234	1,863	0.214	732	0.294
	others (combined)	4,583	0.0009	1,863	0.0006	732	0.006
PGM_2	PGM_2^1	4,583	0.9996	1,863	0.995	732	0.999
	PGM_2^2	4,583	—	1,863	0.004	732	—
	others (combined)	4,583	0.0004	1,863	0.004	732	0.001
PGM_3	PGM_3^1	1,274	0.737	895	0.370	—	—
	PGM_3^2	1,274	0.263	895	0.630	—	—

TABLE 2.5

Estimates of the percentage incidence in the English population of individuals with various combinations of PGM_1 and PGM_3 phenotypes.

		PGM_1 phenotype			
		PGM_1 1	PGM_1 2-1	PGM_1 2	Total
PGM_3 phenotype	PGM_3 1	31.8	19.4	3.0	54.2
	PGM_3 2-1	22.7	13.9	2.1	38.7
	PGM_3 2	4.0	2.5	0.4	6.9
	Total	58.5	35.8	5.5	99.8

PGM_1 and also the PGM_3 isozymes, but that the individual isozymes determined by any one gene locus (e.g. *a*, *b*, *c* and *d* from PGM_1) are all of the same size. It has also been shown by thermostability studies that the PGM_2 isozymes are significantly more stable than either the PGM_1 or the PGM_3 isozymes. With regard to catalytic activity, the present evidence indicates only that all the isozymes exhibit typical phosphoglucomutase activity. It remains possible that further studies will bring to light significant kinetic differences or other differences in their molecular properties.

The nature of the differences between the separate isozymes determined by individual alleles (e.g. isozymes *a* and *c* determined by PGM_1^1, isozymes *e*, *f* and *g* determined by PGM_2^1 etc.) have not yet been elucidated. Evidence from comparisons of the electrophoretic patterns in blood cells of differing mean age suggests that the relative amounts of the isozymes determined by a single allele vary with the average age of the cell population from which they come (Monn 1969b). In general the proportion of activity due to the most cathodal isozyme in each set appears to be greater in the younger and relatively more immature cell populations. This suggests that the most cathodal isozyme in each case may be the primary form synthesised in the cell, and the more anodal isozymes may be derived from this by secondary structural modifications of the protein.

2.5. *Genes and isozymes*

Lactate dehydrogenase and phosphoglucomutase illustrate how in the cells of a single individual a number of distinct and separable proteins may exhibit the same or very similar enzyme activities. Such multiple molecular forms of an enzyme occurring in a single organism are usually referred to as isozymes (or isoenzymes), a term first introduced by Markert and Møller in 1959 and now widely found in the biochemical literature. It is a convenient term operationally, because it implies no specific type of structural relationship between the several protein species which may be observed to have similar enzyme activities. Indeed, as more and more examples of the phenomenon have been studied, it has become clear that different sorts of molecular relationship are likely to be involved in different cases, and they may be brought about in a variety of ways.

The rapid discovery of many examples of isozyme systems in recent years has largely been a product of the extensive use of relatively simple methods of zone electrophoresis in such supporting media as starch gel or acrylamide, combined with the development of sensitive and specific staining methods for

demonstrating the particular zones of enzyme activity in what is usually a very complex protein mixture. This general procedure first developed for esterases (Hunter and Markert 1957) and dehydrogenases (Markert and Møller 1959) has since been extended to many types of enzymes, and has proved to be a peculiarly powerful experimental tool. Often it can be applied directly to crude homogenates of fresh tissues, so that possible *in vitro* changes of the enzyme components of the mixtures can be minimised. The procedure can of course also be applied at successive stages of the separation and purification of the isozyme components so that any alterations produced by particular biochemical manipulations may be readily assessed. Furthermore only relatively small quantities of material are generally required, and this is of particular value in genetical studies where it is usually necessary to examine the isozymes in a given tissue from a large number of different individuals.

The widespread application of this general procedure, combined of course in different cases with other techniques, has led to the identification of a considerable variety of isozyme systems in many different species, and the subject has been extensively reviewed (see, for example, 'Multiple molecular forms of enzymes', *Annals New York Acad. Sci. 151*, 1–689, 1968). It has become clear that the phenomenon is a very general one and quite probably it is exhibited by most enzymes at least in some degree. Furthermore, many non-enzymic proteins have been found to show essentially the same kind of phenomenon, so that it seems that a multiplicity of molecular forms with similar functional activities is a regular feature of most enzymes and proteins.

But such multiple molecular forms of a given sort of enzyme or protein can evidently be generated in a variety of different ways, and with any particular enzyme or protein more than one type of cause may be operating. These different possible causes of isozymes or, more generally, of the multiple molecular forms of functionally similar proteins can be conveniently classified (Harris 1969) into the three distinct categories considered below.

(a) Multiple gene loci coding structurally distinct polypeptide chains of a protein. There may be two or more separate gene loci involved in determining the structures of a set of isozymes or of functionally similar non-enzymic proteins; for example, three distinct loci ('*A*', '*B*' and '*C*') are concerned in specifying the various molecular forms of lactate dehydrogenase and similarly there are at least three distinct loci (PGM_1, PGM_2 and PGM_3) which determine the structures of the various phosphoglucomutase isozymes.

In the case of haemoglobin at least six distinct loci are probably involved ('α', 'β', two 'γ's, 'δ', and 'ε') – see pp. 37–38.

The separate gene loci in each case code for the aminoacid sequences of distinctive polypeptide chains. In some cases the several polypeptides may be associated together in the various members of the set of isozymes, so that the various isozymes differ in the particular combinations of polypeptides they possess. This is so for example in the case of the five standard lactate dehydrogenase isozymes (A_4, A_3B, A_2B_2, AB_3, and B_4), and also of course in the various molecular forms of haemoglobin (Hb A = $\alpha_2\beta_2$, Hb F = $\alpha_2\gamma_2$, etc.). In other cases, however, the polypeptide products of the different loci may separately form the various members of the set of isozymes, and apparently do not occur together in the same protein molecule. This is probably the case with phosphoglucomutase, and also with the lactate dehydrogenase C type isozyme.

(b) Multiple alleles at a single locus. At any given gene locus a number of different alleles may occur in a population of individuals, each allele coding for a structurally distinct version of the particular polypeptide chain, so that the primary structure of the enzyme or protein involved will differ from one individual to another, according to the alleles that they happen to carry at the locus in question. Heterozygotes, since they carry two different alleles, may in general be expected to show a more complex pattern of isozymes than homozygotes. In some cases, the pattern of isozymes in a heterozygote will represent a simple mixture of those occurring in the two corresponding homozygous types. But in other cases, where the enzyme is polymeric, additional 'hybrid' isozymes not present in either type of homozygote may occur (see pp. 33–35).

The extent to which the several polypeptide products of different gene loci which may be concerned in determining a set of isozymes differ in their primary aminoacid sequences no doubt varies very widely from case to case. But it is perhaps not unusual for quite extensive differences to be present. By contrast the different polypeptide products due to allelic differences at a particular locus probably differ from each other in most cases by only single aminoacid substitutions. Thus, in general, isozyme differences in a single individual arising because of the occurrence of multiple gene loci, are likely as a rule to involve a greater degree of molecular difference than isozymes arising because of multiple allelism at a single locus.

However, the critical point of distinction between these separate genetical causes of isozyme formation is that multiple allelism results in differences

between individual members of a population in the pattern of isozymes they form, whereas multiple loci will in general be common to all members of the species, and thus define the overall pattern of isozyme formation that occurs.

(c) 'Secondary' modifications of protein structures. Multiple gene loci and multiple allelism at single loci provide, as it were, the basic genetical framework which defines the main characteristics of the various molecular forms of functionally similar enzymes or proteins which occur in individual members of a species. But the complexity of many isozyme systems which have been studied cannot be fully explained in these terms alone. It seems that secondary modifications of protein structures, subsequent to their primary synthesis of their polypeptide chains on the mRNA templates in the ribosomes, is also an important and general cause of the multiplicity of separable components which are often observed when particular enzymes or proteins are studied by such techniques as electrophoresis.

It is likely, for example, that the characteristic sets of isozymes, each apparently determined by a single allele at one of the phosphoglucomutase loci (fig. 2.17, p. 51), are the consequences of such secondary modifications of the primary protein structure. Similar characteristic sets of isozymes, the several members of which all appear to be the products of a single allele, have been observed in studies on the allelic variants of several other enzymes [e.g. adenylate kinase (Fildes and Harris 1966), adenosine deaminase (Hopkinson and Harris 1968), and peptidase B (Lewis and Harris 1967)], and it is probable that these also occur because secondary modifications of the primary protein products have led to the appearance of multiple forms.

In the case of lactate dehydrogenase, although the main isozyme components observed in various tissues can be simply accounted for in terms of the different tetrameric combinations of the single polypeptide products of the alleles at the three gene loci, it has been found that minor additional components can usually be demonstrated by electrophoresis (Markert 1968). These appear to be consistently associated with the main components observed in any particular tissue, although they may vary somewhat in different tissues. Similar minor components associated with the main electrophoretic components of a number of other enzymes and proteins have also been demonstrated, and they appear to be a not unusual phenomenon. They have been observed both with polymeric enzymes or proteins, such as placental alkaline phosphatase (Robson and Harris 1967) and phosphohexoseisomerase (Detter et al. 1968), and also with monomers such as myoglobin (Epstein and

Schechter 1968). Again, they are probably attributable to secondary modifications of the primary protein products formed.

A wide variety of different structural changes can occur following the primary synthesis of a protein. They may involve, for example, the removal of amide groups from glutamine or asparagine residues in the polypeptide sequences, the oxidation of sulphydryls, the addition of phosphate groups, the addition of carbohydrate groupings of various structures, the cleavage of a polypeptide chain by proteolytic enzymes with the loss of part of the aminoacid sequence, and so on. If such 'secondary' modifications affect some, but not all, of the protein molecules, or affect them in different degrees, or represent a series of steps in a process through which all of them pass, then a characteristic set of multiple molecular forms of the particular protein may be consistently found. Often, indeed, several of the isozymes that are observed in particular cases may simply reflect successive early stages in the process by which an enzyme protein is progressively denatured and broken down *in vivo*.

There are also a number of other possible causes for the appearance of multiple molecular forms of a particular enzyme or protein all coded by the same allele or alleles. Thus in some cases the several forms may have the same primary structures but differ in their three-dimensional conformations. That is to say, they may represent a series of conformational isomers (Kitto et al. 1966), each relatively stable and each, because of the degree to which particular groups are exposed or concealed in their three dimensional arrangements, differing from the others in particular properties such as electrophoretic mobility. In other cases the series of different molecular forms observed may simply represent monomers, dimers, trimers, etc., of a single basic subunit which itself may contain more than one polypeptide chain. An example of this is the series of components seen in the common haptoglobin types Hp 2–1 and Hp 2–2 and discussed on pp. 68–69. Finally, it should be noted that if an enzyme or protein happens to be firmly bound to other types of molecule (large or small) in the intracellular milieu, then what may seem to be multiple molecular forms may be demonstrable when tissue extracts are subjected to a separation technique such as electrophoresis.

2.6. *Chromosomal relations of gene loci determining multiple molecular forms of a particular protein*

2.6.1. Linkage and recombination

When more than one gene locus is concerned in determining the structure of a particular protein or enzyme it is of interest to know something about the relative positions of these loci on the chromosomes.

If we consider two loci, at one of which alleles *A* and *a* may occur, and at the other of which alleles *B* and *b* may occur, then the genotype of an individual heterozygous at both loci will be *Aa,Bb*. Such an individual can form four types of gamete (sperm or ova). These will be *AB*, *Ab*, *aB* and *ab*, and their relative proportions will depend on whether the two loci are on the same or on different chromosomes; and if they are on the same chromosome whether they are close together or relatively far apart. If they are on different chromosomes then the four sorts of gamete will in general occur equally frequently. If they are on the same chromosome then their distance apart determines the extent to which particular genes at the two loci are likely to undergo recombination as a result of crossing-over, a process which occurs during meiosis prior to gamete formation.

Crossing-over involves the exchange of segments between two chromosomes (or more correctly chromatids) when they pair at meiosis (fig. 2.18). It may in general occur anywhere along the length of the chromosomes so that the chance that a cross-over will take place between two particular loci will depend on the distance they are apart. If for example the doubly heterozygous individual *Aa,Bb*, carries the genes *A* and *B* on one chromosome of the pair and a and b on the other, then if a single cross-over occurs between the two loci, recombination of the genes results because one of the chromosomes will now carry *A* and *b*, while the other carries *a* and *B*. Should the two loci be close together on the chromosome then a cross-over between them will be a relatively infrequent event, so the majority of gametes formed will carry a chromosome with one or other of the original combinations *AB* or *ab*, and only a small proportion will have a recombinant, *Ab* or *aB*. Should the loci be well separated on the chromosome then the chance of crossing-over between them is enhanced and so the proportion of recombinants among the gametes correspondingly increased. If the two loci are so far apart that crossing-over occurs freely between them, the situation is then essentially the same as when the two loci are located on different chromosomes. On

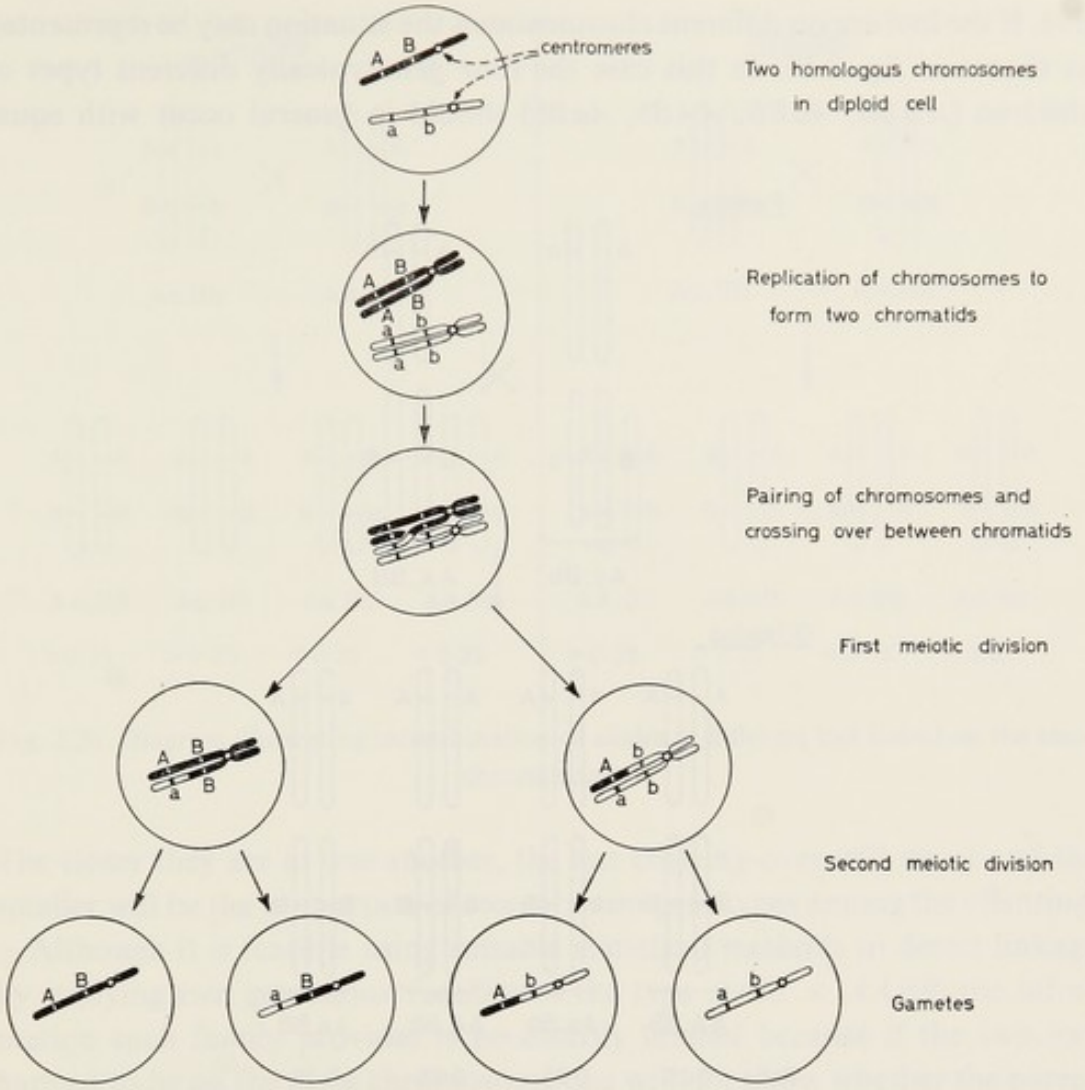


Fig. 2.18. Diagram illustrating crossing over at meiosis.

average 50% of the gametes will carry the original combinations and 50% will be recombinants.

Since recombination between genes at loci on different chromosomes occurs freely, the demonstration of a reduced incidence of recombination between genes at two particular loci implies that these loci are on the same chromosome. Such loci are said to be 'linked', and the detection of linkage involves studying families in which an estimate of the extent of recombination between genes at the two loci can be obtained.

In general the linkage relations of any two loci may be assessed by determining the genotypes of children from matings where one of the parents is heterozygous at both loci. Matings where one parent is heterozygous at both loci (*AaBb*) and the other is homozygous (*AABB*) are particularly informa-

tive. If the loci are on different chromosomes the situation may be represented as shown in fig. 2.19. In this case the four genotypically different types of children (*AABB*, *AaBB*, *AABb*, *AaBb*) should in general occur with equal

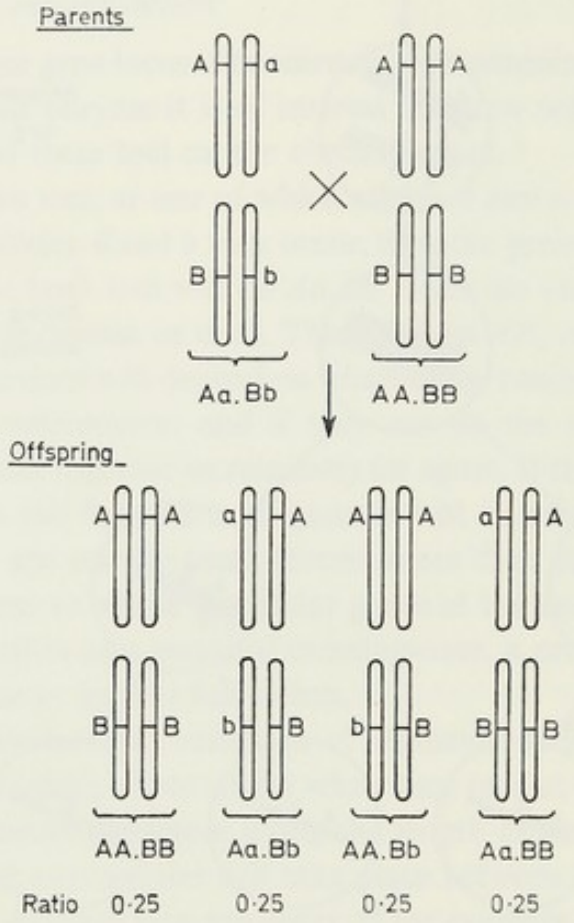


Fig. 2.19. Diagram illustrating recombination of alleles at different loci on separate chromosomes.

probability. On average half the children will have one or other of the parental genotypes, and half the recombinant genotypes. If the loci are on the same chromosome two situations must be considered (fig. 2.20). In the first the heterozygous parent (*AaBb*) has genes *A* and *B* on one member of the pair of homologous chromosomes, and genes *a* and *b* on the other. In the second the heterozygous parent has genes *A* and *b* on one of the pair of chromosomes and genes *a* and *B* on the other. Two genes such as *A* and *B* are said to be in 'coupling' when they are on the same member of the chromosome pair, and in 'repulsion' when they are not. As in the previous case, four different genotypes may occur among the children of such matings, but their proportions will differ according to how far apart the two loci are on the chromosome.

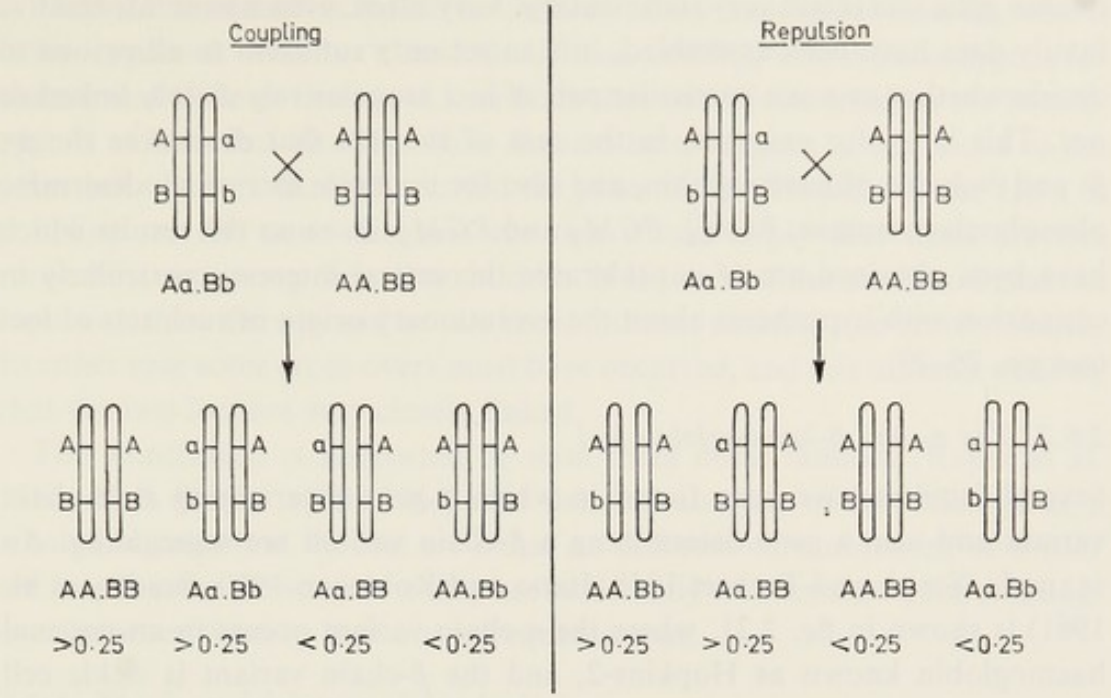


Fig. 2.20. Diagram illustrating recombination of alleles at different loci linked on the same chromosome.

The closer they are to one another, the less crossing-over will occur and the smaller will be the proportion of recombinant genotypes among the offspring.

Although it is feasible using suitable statistical methods to detect linkage by studying two generation families of the type *AaBb* × *AABB*, the information each family provides is necessarily limited because if the two loci happen to be on the same chromosome one will not know whether the parent heterozygous at the two loci carries a particular pair of genes (e.g. *A* and *B*) in coupling or repulsion. This difficulty is best met if data on the grandparents can also be obtained because this can provide information about coupling and repulsion. Data on other relatives may also be informative, and in general the more extensively the pedigree can be investigated, the more informative will be the data for linkage analysis.

In principle it is possible by studying a sufficient number of informative families and by applying the appropriate statistical methods to decide whether any two loci are linked, and if they are, to estimate their distance apart by determining the frequency of crossing over between them (Renwick 1969). From this one could go on to construct linkage maps indicating the relative positions on particular chromosomes of different gene loci. In practice this has proved extremely difficult to do in man for a variety of practical reasons, and consequently our knowledge of the linkage relations of

human gene loci is still very rudimentary. Very often, even where informative family data have been assembled, it is as yet only sufficient to allow one to decide whether or not a particular pair of loci are relatively closely linked or not. This is so, for example, in the case of the loci that determine the α-, β- and δ-chains of haemoglobin, and also for the three loci which determine phosphoglucomutase, PGM_1, PGM_2 and PGM_3. Even so the results which have been obtained are of considerable theoretical interest, particularly in connexion with hypotheses about the evolutionary origin of such sets of loci (see pp. 76–79).

2.6.2. The α- and β-haemoglobin loci

Several families have been found in which a gene determining an α-chain variant and also a gene determining a β-chain variant are segregating. An example (Smith and Torbert 1958, Itano and Robinson 1960, Bradley et al. 1961) is shown in fig. 2.21, where the α-chain variant occurs in an unusual haemoglobin known as Hopkins-2, and the β-chain variant is sickle cell

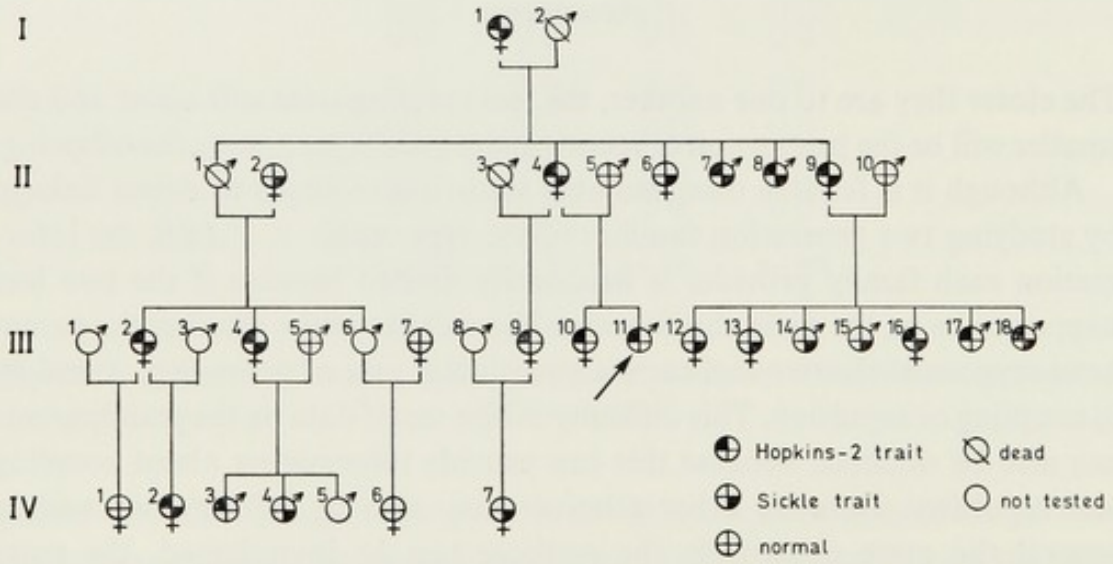


Fig. 2.21. Pedigree showing segregation of alleles determining the α-chain haemoglobin variant Hopkins-2, and the β-chain variant Hb S (Smith and Torbert 1958, Bradley et al. 1961.)

haemoglobin. Several individuals in the pedigree are heterozygous for the gene determining Hopkins-2 and its allele which determines the normal α-chain, and also for the sickle-cell gene and its allele which determines the normal β-chain. The segregation of these genes in the pedigree suggests that the two loci determining the α- and β-chains cannot be very closely linked.

Thus if one considers the seven children of the mating between II_9 who is heterozygous at both the 'α' and the 'β' loci, and II_{10} who is homozygous for the normal allele at these loci, one finds that two of the children are heterozygous at both loci like II_9, while five are heterozygous at one locus or the other but not at both. If II_9 carried the genes for Hb Hopkins-2 and Hb S in coupling on the same chromosome this would imply that there are five recombinants among the seven children. If II_9 carried the abnormal genes in repulsion then there would be two recombinants and five non-recombinants. In either case some cross-overs must have occurred, and this makes it unlikely that the two loci are very closely linked.

This conclusion is supported by data from other families (Raper et al. 1960, Hall Craggs et al. 1964) and detailed analysis of the findings as a whole indicates that the loci are either on different chromosomes, or if they are on the same chromosome they are sufficiently well separated for crossing-over between them to be a not infrequent event.

2.6.3. The β- and δ-haemoglobin loci

A number of families have also been found which enabled one to examine the linkage relation between the locus determining the β-chain and the locus determining the δ-chain. Here a quite different result has been obtained. No definite example of a cross-over has yet been observed, though more than sixty children of matings where one parent is doubly heterozygous have been studied (Mishu and Nance 1969).

A typical pedigree (Horton and Huisman 1963) is shown in fig. 2.22.

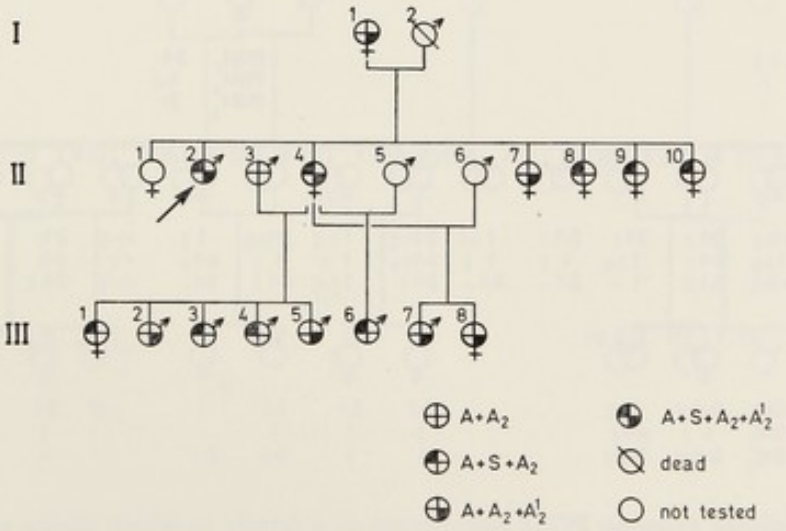


Fig. 2.22. Pedigree showing segregation of alleles determining the β-chain variant Hb S and the δ-chain variant Hb A_2^1 (Horton and Huisman 1963).

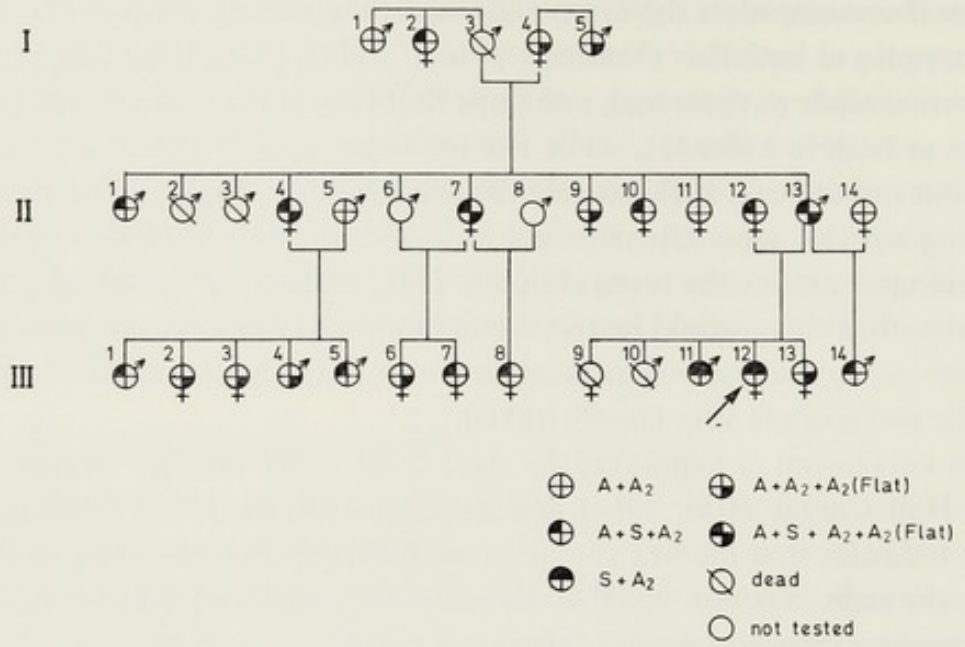


Fig. 2.23. Pedigree showing segregation of alleles determining the haemoglobin β-chain variant Hb S and the δ-chain variant Hb A_2 Flatbush (Ranney et al. 1963).

Here II_4 is heterozygous at the β-locus for the sickle-cell gene, and is also heterozygous at the δ-locus for the gene which determines an Hb A_2 variant referred to as Hb A_2^1. Of her eight children by three fathers, four show only the sickle-cell trait (i.e. are heterozygous for the β-locus, but not for the

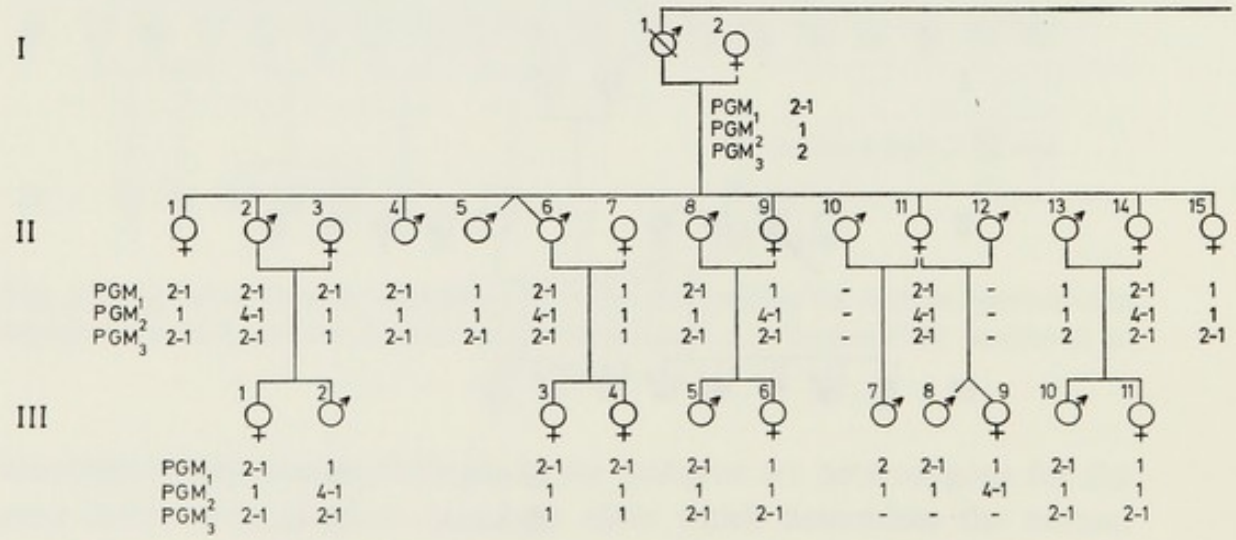
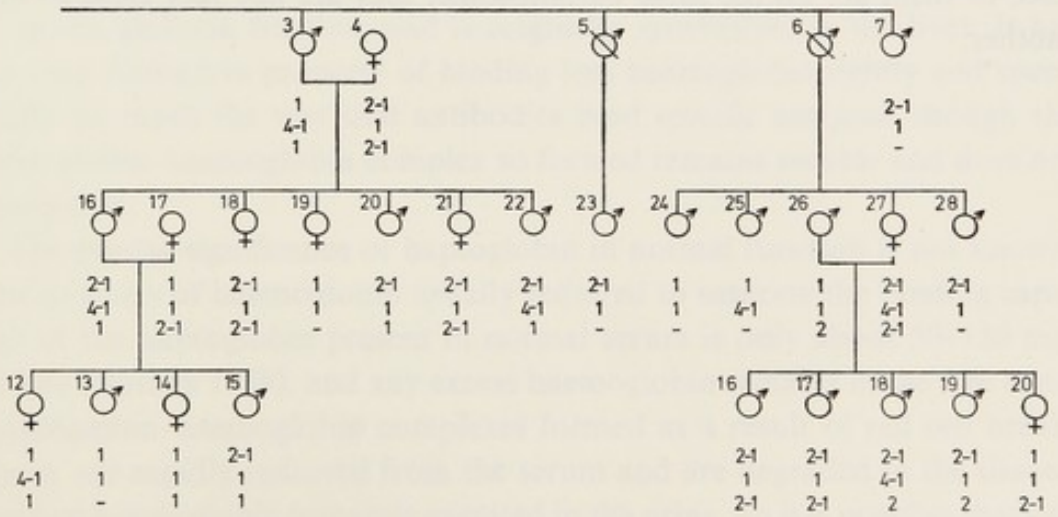


Fig. 2.24. Pedigree showing PGM phenotypes in a family in which alleles at each of the three phosphoglucomutase loci *PGM*$_1$, *PGM*$_2$ and *PGM*$_3$ are segregating (Parrington et al. 1968).

δ-locus), and four show only the Hb A_2^1 trait (i.e. are heterozygous at the δ-locus, but not the β-locus). There were no children who were either homozygous or heterozygous at both loci. The pedigree thus supports the idea that the two loci are linked and suggests that II_4 carried the β^S- and the δ^1-genes in repulsion. She presumably received the β^S-gene from her father and the δ^1-gene from her mother. Fig. 2.23 shows another pedigree (Ranney et al. 1963) illustrating the same phenomenon, but in this case with a different mutant allele at the δ-locus. Taken together these and the other pedigrees of families where mutant alleles at both the β-locus and the δ-locus were found to be segregating provide convincing evidence that these two loci are very closely linked (see also Lepore haemoglobins pp. 80–83).

So far there is no data available bearing on the linkage relations of the locus determining the γ-chains of haemoglobin and the loci determining the α- or the β-chains. This is not surprising in view of the difficulty of studying the segregation in families of variants of Hb F which can usually only be detected in foetal life or in the newborn.

At present, then, one may say that the loci determining the β- and δ-polypeptide chains of haemoglobin lie close together on the same chromosome, possibly in immediate juxtaposition, and that the locus determining the α-polypeptide chain lies either on a quite different chromosome or, if it is on the same chromosome as the β- and δ-loci, it must be some distance away from them.



2.6.4. *The phosphoglucomutase loci: PGM_1, PGM_2 and PGM_3*

Studies of the segregation pattern in families in which there were parents who were heterozygous for alleles at two and in some cases all three of the phosphoglucomutase loci, indicate the absence of close linkage.

A particularly informative pedigree (Parrington et al. 1968) is shown in fig. 2.24. The propositus (II_6) was an individual who was found in the course of routine studies on the enzyme in his red cells to be heterozygous for a rare allele at the PGM_2 locus (phenotype PGM_2 4–1), as well as being heterozygous for the two common alleles at the PGM_1 locus (phenotype PGM_1 2–1). The PGM_3 phenotype cannot be routinely determined in red cells, but by examining fibroblasts grown in tissue culture from a small skin biopsy, it was shown that the propositus was also heterozygous at the PGM_3 locus (phenotype PGM_3 2–1). Thus he was heterozygous at all three loci. By similar studies on red cells and on fibroblasts grown in tissue culture, from more than forty of his relatives the segregation pattern shown in the pedigree (fig. 2.24) was determined. Inspection of the pedigree shows that recombination between PGM_1 and PGM_2, PGM_1 and PGM_3, and PGM_2 and PGM_3, must be relatively frequent, and this has been confirmed by more exact linkage calculations.

Evidence against close linkage of PGM_1 and PGM_2, and against close linkage of PGM_1 and PGM_3 has also been obtained in other studies (Hopkinson and Harris 1965, 1968). Thus one may conclude that the loci PGM_1, PGM_2 and PGM_3 either occur on different chromosomes, or if two or all three of them are on the same chromosome, they are not very close to one another.

CHAPTER 3

Duplications and deletions, and their effects on protein structure

3.1. The haptoglobin variants

A gene mutation which results in a single aminoacid substitution in a protein generally represents, as we have seen, a single base change in the DNA sequence of the gene. But other kinds of mutational events, which may involve changes such as duplications or deletions of particular stretches of the DNA sequence can also occur. Like the single base changes, these alterations in the sequence will be perpetuated by the ordinary processes of DNA replication. They will also be reflected by corresponding alterations in the structure of the polypeptide or polypeptides coded by the gene or genes affected by the change. An example of an alteration in protein structure that has apparently been brought about by a mutational event of this sort is provided by one of the common variant forms of the serum protein, haptoglobin.

Blood serum contains a complex mixture of many different proteins which originate in various tissues. Haptoglobin constitutes part of the so-called α_2 serum globulin fraction, and is originally synthesised in the liver. It has the very distinctive property of binding free haemoglobin tightly and specifically in much the way that antibodies bind specific antigens, though the haptoglobin–haemoglobin complex so formed remains soluble and does not precipitate.

The precise significance of haptoglobin in normal function is not known. The quantity of haemoglobin usually required to saturate the binding capacity of the haptoglobin present in normal serum is only about 50–150 mg/100 ml (Nyman 1959), and any excess haemoglobin remains in the free state. Haptoglobin–haemoglobin complexes formed as a result of red cell breakdown, are rapidly removed from the serum and are degraded in the tissues, but free haemoglobin is mainly excreted in the urine. So it is possible that one function of haptoglobin is to minimise the loss of iron by the body. Haptoglobin may also be significant in bile pigment formation because it has been shown that while the haem in native haemoglobin is resistant to the attack of

a liver enzyme, known as α-methenyl oxygenase, which converts haem to a precursor of biliverdin, haem in the haptoglobin–haemoglobin complex is readily attacked by this enzyme (Nakajima et al. 1963).

Smithies (1955) discovered that when serum from different individuals is examined by starch gel electrophoresis several distinct and quite characteristic haptoglobin types can be recognised (fig. 3.1). The most commonly

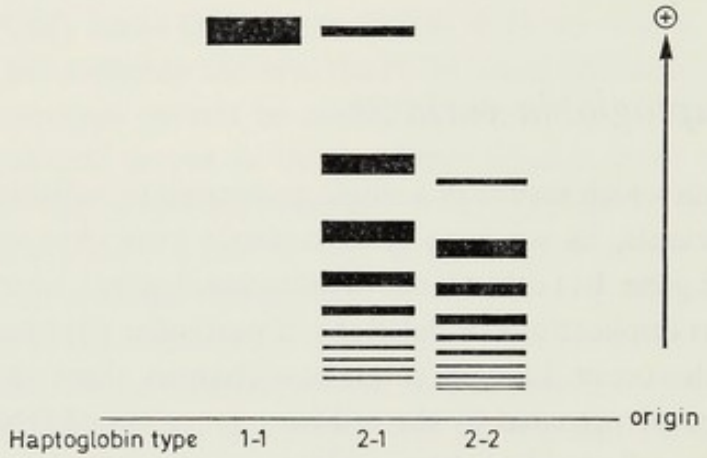


Fig. 3.1. The three common haptoglobin types: Hp 1-1, Hp 2-1 and Hp 2-2. Diagram of patterns of components after electrophoresis in starch gel (*p*H 8.6).

occurring types are called Hp 1-1, Hp 2-1 and Hp 2-2, and they differ from each other both in the electrophoretic mobilities and also in the number of protein components present. In type 1-1 only a single haptoglobin component is seen. A component with the same mobility can be detected in type 2-1 but not in type 2-2. However both the 2-1 and 2-2 types show a whole series of other haptoglobin components migrating more slowly than the 1-1 component, and each of these components in type 2-1 has a different mobility from those in type 2-2. Thus the pattern of components is qualitatively distinct in the three types, and simple mixing for example of 1-1 and 2-2 sera will not produce a 2-1 pattern.

In European populations virtually all individuals can be classified into one or other of these three types; about 16% are type 1-1, about 48% are type 2-1 and about 36% are type 2-2. The types are genetically determined and the initial family studies led to the suggestion that a pair of alleles at an autosomal locus were involved (Smithies and Walker 1955, 1956). These were called Hp^1 and Hp^2, and it was supposed that Hp 1-1 individuals were homozygous for Hp^1, Hp 2-2 individuals homozygous for Hp^2 and Hp 2-1 individuals were heterozygous Hp^1Hp^2. It appeared therefore that the homozygotes differed from one another both in the mobility and the number

of protein components they formed, and also that the heterozygotes formed a series of protein components qualitatively different from those present in either type of homozygote.

Structural studies on the purified haptoglobins from sera of the different types carried the matter much further. It turned out that the multiple components characteristic of types 2-1 and 2-2 represent a series of polymers of increasing molecular weight (Smithies and Connell 1959). Both they and the single component in type 1-1 each contain two sorts of non-identical polypeptide chains, which are represented two or more times in a haptoglobin molecule according to its molecular size. The two sorts of polypeptide are called the α- and β-chains, and the differences between the three common haptoglobin types depend on structural differences in the α-chain (Smithies et al. 1966). The β-chains are evidently the same in each type (Cleve et al. 1967). In the haptoglobin molecule the chains are apparently cross-linked by disulphide bonds.

Electrophoretic studies on the separated α-chains from the different types revealed further genetically determined heterogeneity (Connell et al. 1962, Smithies et al. 1962a, 1966). Two sorts of α-chain can be obtained from preparations of type 1-1 haptoglobin. These are called hp1Fα and hp1Sα, the former having a somewhat faster electrophoretic mobility than the latter

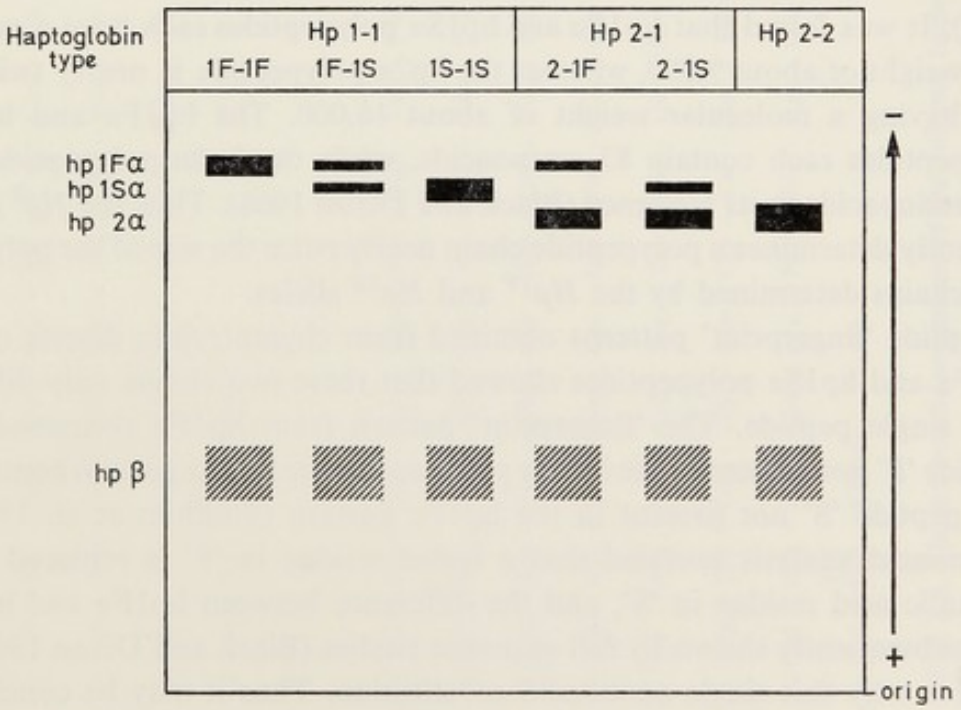


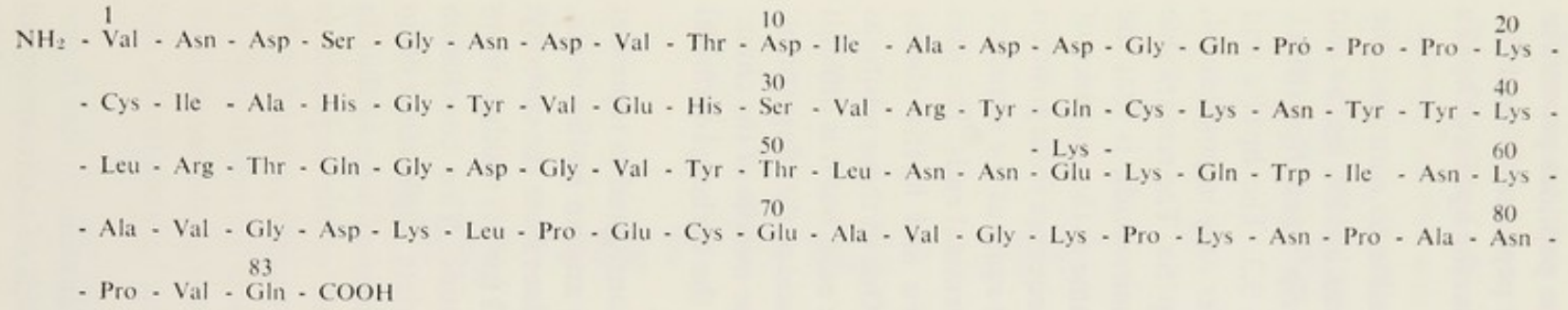
Fig. 3.2. Electrophoretic separation of hp α-chains in different haptoglobin types. Starch gel electrophoresis in formate buffer (pH 4.0) and 8.0 M urea (Smithies et al. 1962a, 1966).

under the particular conditions used for their separation (fig. 3.2). Type 1-1 haptoglobin from some individuals contains only hp1Fα, that from other 1-1 individuals contains only hp1Sα, and there is a third group of type 1-1 individuals in which both hp1Fα and hp1Sα are present. These 'concealed' differences within the Hp 1-1 type are genetically determined and their manner of segregation within families shows that two sorts of Hp^1 allele must exist. These are designated Hp^{1F} and Hp^{1S}. Homozygotes for one of these alleles form a type 1-1 haptoglobin containing only one kind of α-chain. Heterozygotes ($Hp^{1F}Hp^{1S}$) form type 1-1 haptoglobin which contains molecules with both kinds of α-chain.

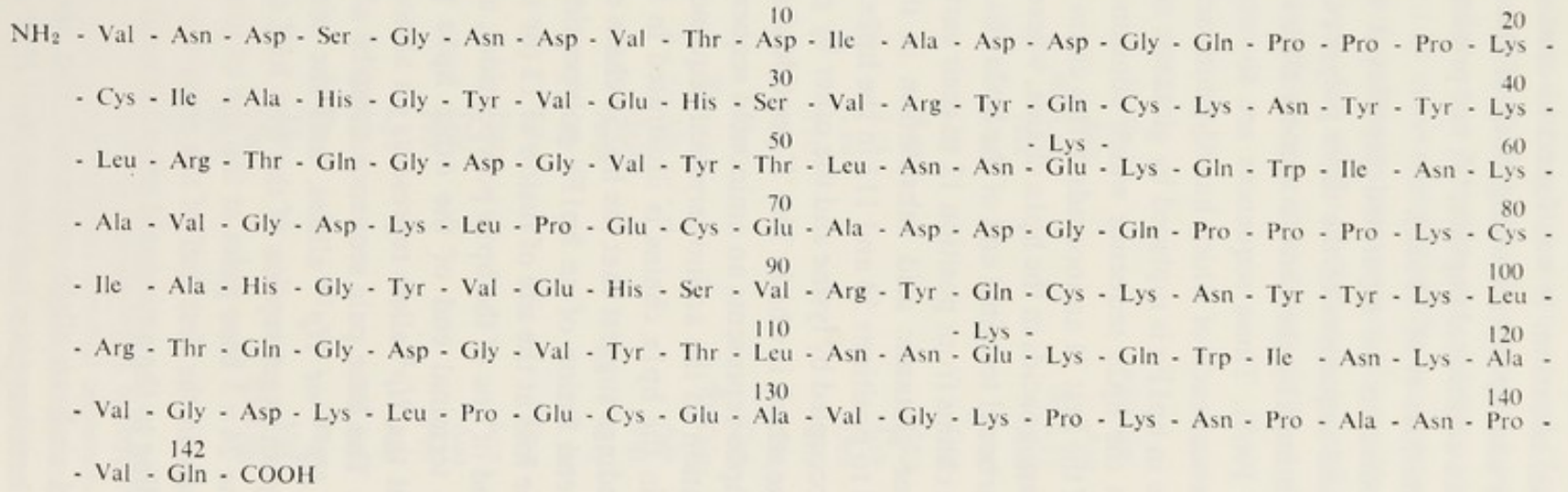
The α-chain obtained in most cases from type 2-2 haptoglobin migrates electrophoretically as a single slower-moving zone. It is referred to as hp2α. As would be expected two subtypes of type 2-1 can be distinguished, one corresponding to the genotype $Hp^{1F}Hp^2$ contains the α-chains hp1Fα and hp2α, the other from the genotype $Hp^{1S}Hp^2$ contains hp1Sα and hp2α. It is of interest that the three Hp 1-1 types cannot be clearly distinguished from one another by routine electrophoresis of the native haptoglobins, nor can the two different types of Hp 2-1.

Studies on the structure of the isolated α polypeptides revealed remarkable and at the time quite unexpected differences between hp1Fα and hp1Sα on the one hand, and hp2α on the other (Smithies et al. 1962b, Connell et al. 1966). It was found that hp1Fα and hp1Sα polypeptides each have a molecular weight of about 9,000, whereas the hp2α polypeptide is nearly twice as big, having a molecular weight of about 16,000. The hp1Fα and hp1Sα polypeptides each contain 83 aminoacids, while the hp2α polypeptide has 142 aminoacids in its sequence (Black and Dixon 1968). Thus the Hp^2 allele evidently determines a polypeptide chain nearly twice the size of the polypeptide chains determined by the Hp^{1F} and Hp^{1S} alleles.

Peptide 'fingerprint' patterns obtained from chymotryptic digests of the hp1Fα and hp1Sα polypeptides showed that these two chains only differed in a single peptide. The 'fingerprint' pattern from hp1Fα contained one peptide 'F' not present in the hp1Sα pattern, and the hp1Sα pattern contained one peptide 'S' not present in the hp1Fα pattern (Smithies et al. 1962b). Aminoacid analysis revealed that a lysine residue in 'F' is replaced by a glutamic acid residue in 'S', and the difference between hp1Fα and hp1Sα was subsequently shown by full sequence studies (Black and Dixon 1968) to involve only this single aminoacid substitution. Thus it may be concluded that the alleles Hp^{1F} and Hp^{1S} have arisen one from the other by a mutation of the standard kind, involving only a single base change in the DNA.



(a)



(b)

Fig. 3.3. The aminoacid sequences of the different haptoglobin α-chains hp1Fα, hp1Sα and hp2α (Black and Dixon 1968). (a) hp1Fα: Lys at position 54; hp1Sα: Glu at position 54. (b) hp2α: because the order of the F/S substitutions is not known both Lys and Glu are shown at positions 54 and 113. (Aminoacid abbreviations as in caption to fig. 1.5, p. 13.)

The peptide 'fingerprint' pattern obtained from the hp2α chain revealed a quite unusual situation. It was found to contain all the peptides present in both hp1Fα and hp1Sα including both of the specific peptides 'F' and 'S'. It also had an additional peptide, 'J', not present in either the hp1Fα or hp1Sα patterns (Smithies et al. 1962b).

From studies on the separated peptides it was possible to construct the full aminoacid sequences of the three polypeptide chains hp1Fα, hp1Sα and Hp2α determined by the three common alleles Hp^{1F}, Hp^{1S} and Hp^{2} (Black and Dixon 1968). These sequences are shown in fig. 3.3. The hp1Fα and hp1Sα sequences each of which has 83 aminoacids, differ only in position 54, where lysine in hp1Fα is replaced by glutamic acid in hp1Sα. The remarkable thing about the hp2α sequence which contains 142 aminoacids, is that the sequence of the first 71 aminoacids in this chain is the same as the sequence of the first 71 aminoacids in the hp1α chains, whereas the sequence from position 71 to the carboxyl terminal end of the hp2α chain is the same as the sequence in the hp1α chains from position 12 to their carboxyl terminal ends. The sites of the lysine–glutamic acid substitution in the hp1Fα and hp1Sα chains correspond to positions 54 and 113 in the hp2α chain. One of these positions in hp2α is occupied by lysine and the other by glutamic acid, but it is difficult to determine which is which. The 'J' peptide characteristic of the 'fingerprint' pattern of hp2α represents an aminoacid sequence in the middle of the chain, and its relation to the amino-terminal sequences and the carboxyl-terminal sequences in the hp1α chains is illustrated in fig. 3.4.

These findings suggest that the hp2Sα chain can be considered to represent an end to end fusion of an hp1Fα polypeptide chain and an hp1Sα chain, but with the loss at the site of fusion of 12 (or 13) residues from the carboxyl terminal end of one of the hp1α polypeptides, and of 12 (or 11) residues from the amino terminal end of the other hp1α polypeptide. This conclusion implies that the Hp^{2} allele represents an almost complete duplication of an Hp^{1} allele. The base pair sequence must be almost twice as long as that in either the Hp^{1F} or Hp^{1S} alleles, and the second half of the sequence is presumably an exact replica of the first half of the sequence except for the loss of some 24×3 base pairs at the site of junction, and the substitution of one base pair in the first half of the sequence by another in the second half corresponding to the aminoacid substitutional difference between hp1Fα and hp1Sα.

How did such an allele originate? The likely explanation is that it first arose in a heterozygous individual $Hp^{1F}Hp^{1S}$ as a result of a chromosomal rearrangement which caused virtually the whole of the base sequences of the

two alleles Hp^{1F} and Hp^{1S} to be brought together end to end on a single chromosome. One way in which this might have occurred (Smithies 1964) is as a consequence of an event in which breaks happened to occur more or less simultaneously in each of the two homologous chromosomes (or chromatids) and this was followed by cross reunion of the broken ends (fig. 3.5). We may suppose that one of these breaks occurred in the region of the distal end of the haptoglobin locus on one of the chromosomes, and the other at the proximal end of the haptoglobin locus in the homologous chromosome.

(a)		(b)		(c)
	|			
	|	58 Ile		58 Ile
	|	Asn		Asn
1 Val	|	60 Lys		60 Lys
Asn	|	Ala		Ala
Asp	|	Val		Val
Ser	|	Gly		Gly
Gly	|	Asp		Asp
Asn	|	Lys		Lys
Asp	|	Leu		Leu
Val	|	Pro		Pro
Thr	|	Glu		Glu
10 Asp	|	Cys		Cys
Ile	|	70 Glu		70 Glu
Ala	|- - - -	Ala - - - -	|	Ala
Asp		Asp	|	Val
Asp		Asp	|	Gly
Gly		Gly	|	Lys
Gln		Gln	|	Pro
Pro		Pro	|	Lys
Pro		Pro	|	Asn
Pro		Pro	|	Pro
20 Lys		Lys	|	Ala
Cys		80 Cys	|	80 Asn
Ile		Ile	|	Pro
Ala		Ala	|	Val
His		His	|	83 Gln
Gly		Gly	|	
Tyr		85 Tyr	|	

Fig. 3.4. Relationship of sequence of 'J' (junction) peptide found in chymotryptic digest of hp2α polypeptide chain, to the amino and carboxyl terminal sequences of the hp1α chains. (a) Amino-terminal sequence of hp1α (residues 1-26). (b) 'Junction' peptide obtained from chymotryptic digest of hp2α (residues 58-85). (c) Carboxyl-terminal sequence of hp1α (residues 58-83).

This was followed by cross-reunion so that one chromosome came to contain the major portions of the two alleles aligned end to end, while the other contained only a fragment of each. Thus two new chromosomes would

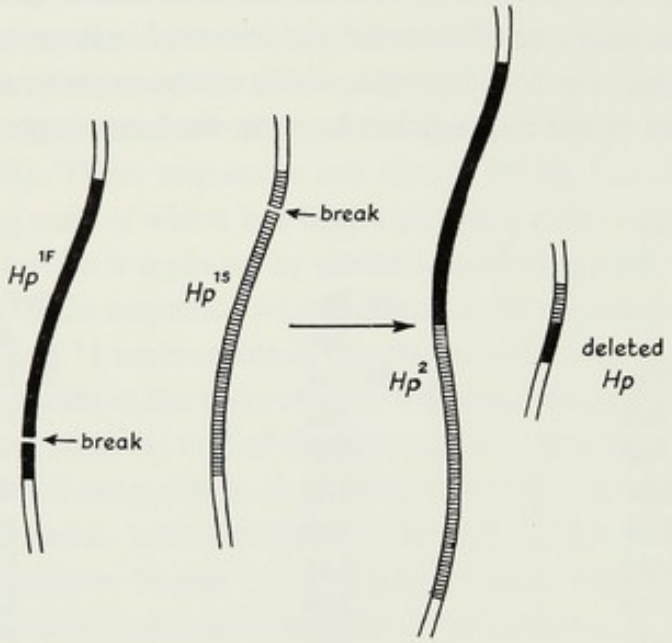


Fig. 3.5. Diagram illustrating postulated formation of Hp^2 allele as a result of breaks in Hp^{1F} and Hp^{1S} alleles followed by aberrant reunion.

appear. One would contain an almost complete duplication of the Hpα locus. This would be the Hp^2 allele. In the other new chromosome the locus would be effectively deleted, except for a short base pair sequence which would probably not define a viable polypeptide chain. Formally the event could be regarded as a reciprocal translocation occurring between two homologous chromosomes.

Translocation of segments between different chromosomes presumably resulting from random breaks followed by aberrant reunions are known to occur, and in population surveys carried out using standard cytogenetical techniques, such chromosomal translocations have been shown to exist in a small but appreciable fraction of the general population (Court Brown and Smith 1969). For the most part these have only been identified when the chromosomal changes produced are sufficiently gross that they can be recognised microscopically either because of an alteration in the dimensions of particular chromosomes observed in mitosis, or because of aberrations of pairing observed in meiosis. One may presume that similar changes not

detectable under the microscope are at least as common if not more frequent. Thus the idea that the Hp^2 gene originally arose in this way is a plausible one.

Another way in which it has been suggested that the Hp^2 allele might have arisen, is as a consequence of what is known as unequal crossing over (Smithies et al. 1962b, Black and Dixon 1968). If in the course of meiosis in a heterozygous individual $Hp^{1F}Hp^{1S}$ there were aberrant pairing of the two homologous chromosomes so that the Hp^{1F} and the Hp^{1S} alleles were misaligned, then should crossing-over occur between the distal end of one allele and the proximal end of the other, the new Hp^2 allele could have been produced on one of the chromosomes (fig. 3.6). The phenomenon of unequal crossing-over (see pp. 79–86) is known to happen but is thought to require close homo-

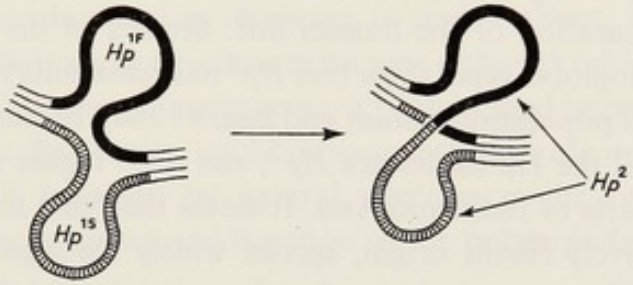


Fig. 3.6. Diagram illustrating postulated formation of Hp^2 allele as a result of unequal crossing over between Hp^{1F} and Hp^{1S} alleles.

logy between the segments of DNA in the two regions of the chromosomes where the crossing-over occurs, so that aberrant pairing is possible. Black and Dixon (1968) arguing from the aminoacid sequences have suggested that there may indeed be sufficient homology between the proximal and distal regions of the DNA coding for hp1Fα and hp1Sα, to make this hypothesis of the origin of Hp^2 reasonable. But although the matter is still unresolved, the hypothesis of random breaks and reunions seems perhaps the most plausible one at the present. An essential difference between the two hypotheses is that the unequal crossover hypothesis requires that the aberrant event took place during pairing at meiosis, while the other hypothesis does not necessarily require this.

Whether or not the original appearance of the Hp^2 allele was a consequence of unequal crossing-over, it seems that this general process is significant in the formation of other haptoglobin variants derived from hp2α. This point is discussed later (pp. 83–86).

3.2. Gene duplication and protein evolution

One apparent consequence of the extended length of the polypeptide Hp2α is that haptoglobin proteins containing it tend to form a series of stable polymers of increasing molecular size. This is probably because the additional cysteine residues on the extended polypeptides can form disulphide bridges between the chains and thus allow the generation of stable haptoglobin molecules with larger numbers of polypeptide subunits. It is of some interest that the multiple polymers characteristic of haptoglobin types 2-2 and 2-1 in man have not been found in comparative studies on a number of primate haptoglobins (Parker and Bearn 1961). This has led to the suggestion that the chromosomal rearrangement giving rise to the duplicated polypeptide structure was probably a single event which occurred some time after the evolutionary separation of the human line. Studies of the world-wide distribution of haptoglobin types show that Hp^2 has remarkably high frequencies in many different populations (Shim and Bearn 1964). In Europe for example about 60% of all the Hp alleles are Hp^2, and even higher frequencies have been found in parts of India and Asia. It seems then that the Hp^2 allele has, despite its relatively recent origin, spread widely through the species (see fig. 8.2, p. 222). Thus one may in fact be witnessing an evolutionary change in the gross structure of a particular protein (Dixon 1966).

In most proteins a polypeptide chain of increased length resulting from such a partial gene duplication is not likely in the great majority of cases to give rise to a stable conformation. The three-dimensional arrangement would generally be profoundly altered, and particularly when the polypeptide is required to fit together with other polypeptides in the protein it seems improbable that a viable and functional protein product would usually emerge. Thus one would expect that only very rarely would what appears to be the essentially random process of breaks and reunions of chromosomes give rise to a viable protein with a grossly altered structure.

Quite apart from this possibility however, chromosomal rearrangements have probably played an important role in protein evolution in a somewhat different way. If the breaks were to occur not in the middle of the DNA sequence defining a particular polypeptide chain, but at the very end of it or at its beginning, and if following two such breaks at appropriate sites in homologous chromosomes (or in a pair of chromatids) cross-reunion should take place, then complete gene duplication could result. In other words one chromosome would come to contain two similar genes aligned sequentially, each capable of determining separately a polypeptide chain of the same

length and with unaltered structure. This sort of duplication could of course involve a longer segment of chromosome containing a number of genes. A variety of other kinds of chromosomal rearrangement due to breaks and aberrant reunions and resulting in gene duplication may also be envisaged (Watts and Watts 1968). They may result in the duplicates being aligned sequentially in the same chromosome as in the haptoglobin case or separated by a non-duplicated region, and the duplicated sequences may run in the same direction or be reversed relative to one another. Furthermore under certain circumstances the duplicates could come to lie in two quite different chromosomes. For instance if two breaks occurred in one chromosome (or its chromatid) and a single break in another the free segment from the one could become inserted between the broken ends of the other with subsequent reunion.

Complete gene duplication, however it comes about, is potentially of considerable evolutionary significance because although initially no structural alteration in the corresponding enzyme or protein need be present, mutations of the single base change type could subsequently occur in one or other of the duplicates and if favoured by natural selection might eventually lead to divergence in the structure and function of the polypeptide products of the two loci.

The occurrence of duplications of genetic material in chromosomes has been recognised for many years, and their likely significance of evolution was appreciated long before their immediate relevance to protein structure was appreciated. Writing in 1951, E.B. Lewis summarised the argument succinctly. 'A gene which mutates to a new function should in general lose its ability to produce its former immediate product, or suffer an impairment in that ability. Since it is unlikely that this old function will usually be an entirely dispensable one from the standpoint of the evolutionary survival of the organism, it follows that the new gene will tend to be lost before it can be tried out, unless, as a result of establishment of a duplication the old gene has been retained to carry out the old function'.

The concept can effectively be illustrated by considering the distinct though structurally similar polypeptide chains, α, β, γ and δ which occur in the different forms of normal haemoglobin; A, A_2 and F. There are considerable homologies in the aminoacid sequences of these four polypeptides, and these can be most simply accounted for by assuming that the genes which determine them were originally derived from a common ancestral gene. The evolutionary process can be envisaged as having involved a successive series of gene duplications followed in each instance by the divergent evolution

of the products formed as a result of point mutations causing different amino-acid substitutions. A simple scheme proposed by Ingram (1961) is illustrated in fig. 3.7. The order in which it is assumed that the duplications occurred to

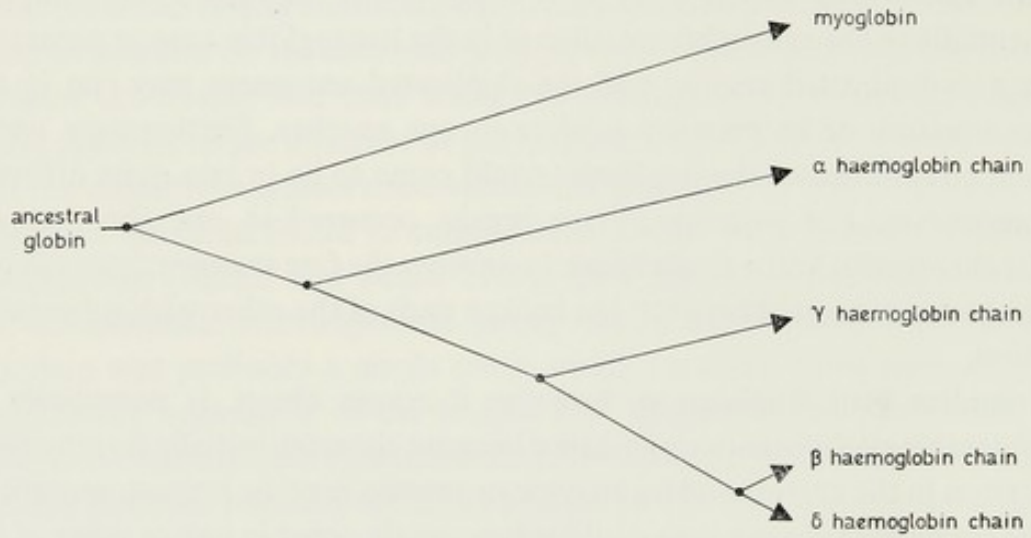


Fig. 3.7. Scheme for the evolution of the α, β, γ and δ haemoglobin polypeptide chains, and also the myoglobin chain from a common ancestral form of globin (after Ingram 1961). Each ● indicates the occurrence of a gene duplication.

give rise to the ancestral forms of the four different loci is consistent with the degree of homology of the aminoacid sequences in the four polypeptide chains as they occur in the human species today. Thus the β- and δ-chains differ in only ten out of 146 positions in their aminoacid sequences and can therefore be regarded as having appeared as separate entities fairly recently in evolutionary history. It is significant that these two loci have been shown by linkage studies to lie very close together on the same chromosome (pp. 63–65), and this of course fits in very well with the idea that they might originally have arisen from a gene duplication. The α polypeptide chain has only 141 aminoacid residues whereas the β-, γ- and δ-chains each have 146 and the correspondences in sequence between the α-chain and the other chains, though striking, are much fewer. This suggests therefore an earlier divergence of the ancestral α-chain from the others. The difference in number of residues can be visualised as having arisen either from small deletions or from accretions of base sequences resulting from chromosome breaks and reunions subsequent to divergence. The finding that the α-locus is not closely linked to the β- or δ-loci presumably implies that the original duplication was associated with separation of the loci or that a translocation of a chromosomal segment

causing the separation of these loci occurred at some subsequent stage in their evolutionary history.

Ingram also suggested that the gene determining myoglobin was also originally evolved from the ancestral gene which was the progenitor of the haemoglobin genes. Myoglobin is a haem protein concerned with oxygen transfer in muscle. Unlike haemoglobin it consists of only a single polypeptide chain, which is however in size (about 150 aminoacid residues) not very different from the polypeptide chains found in haemoglobin. The three-dimensional structure of myoglobin has been shown to have a remarkable similarity to each of the haemoglobin polypeptides, and it also exhibits some degree of homology with their aminoacid sequences. One may imagine that the evolution of haemoglobin chains led at some stage to the possibility of their combining together to form dimers and later tetramers while this did not happen with the myoglobin chain. The tetrameric form allows the possibility of so called haem–haem interaction, and the ultimate development of the characteristic haemoglobin molecule with its sigmoid oxygen dissociation curve specially adapted for the transfer of oxygen from the lungs to the tissues. Myoglobin with its single polypeptide chain and its single haem group gives a hyperbolic oxygen dissociation curve. This is appropriate to the role it plays in providing a kind of oxygen store in muscle.

Although the details of such evolutionary changes are still very far from clear, the underlying idea that structurally different polypeptides which occur in the same or in similar proteins may have originated by duplications from a single ancestral gene, is an important one. It provides a new insight into the structural complexities of many enzymes and other proteins where the occurrence of more than one structurally distinct but nevertheless very similar polypeptide chain is proving to be a not uncommon feature.

3.3. Unequal crossing-over

Crossing-over is a normal phenomenon which takes place between the chromatids of homologous chromosomes when they come together and pair at meiosis. The mechanism by which pairing or synapsis takes place in meiotic cell division is not understood. It is usually a very exact process, and it presumably depends in some way on a precise matching gene for gene along the whole length or at least long stretches of the synapsing chromosomes. This in turn is presumably a reflexion of the close correspondence of the nucleotide sequences in the separate pairs of alleles. As a consequence of this close matching of the pairs of homologous chromosomes, crossing-over,

when it takes place at some particular site along the chromatids, does not normally alter or disrupt the arrangement of the DNA sequences, as might be expected to occur if they were misaligned.

But misalignment or mispairing at synapsis is more likely to happen in regions where a duplication of the DNA sequence exists, and if a cross-over should take place in such a misaligned region, then rearrangements of the DNA sequences and hence alterations in the structures of the polypeptides coded by these sequences may be brought about. This phenomenon is usually referred to as 'unequal crossing-over'. It provides an elegant explanation for the characteristic structural abnormalities found in a particular group of rare haemoglobin variants known as the 'Lepore' haemoglobins (Baglioni 1962a).

3.3.1. The Lepore haemoglobins

The Lepore haemoglobins, so-called from the name of the family in which the defect was first recognised (Gerald and Diamond 1958), contain normal haemoglobin α-chains, but the non-α chains have a quite unusual structure. Like the normal β- and δ-haemoglobin chains, the Lepore non-α chains have 146 aminoacids. But while the first part of the sequence is the same as that of the first part of the sequence of the δ-chain, the remainder of the sequence corresponds to the latter part of the sequence of the normal β-chain. Two sorts of Lepore haemoglobin have been studied in detail. In one of these, known as Hb-Lepore Washington, the change-over from the δ-like sequence to the β-like sequence occurs between positions 87 and 116 (Baglioni 1962a, Labie et al. 1966). In the other, which is known as Hb-Lepore Hollandia, the change-over occurs between positions 22 and 50 (Barnabas and Muller 1962, Curtain 1964).

These Lepore haemoglobins have been mainly found in heterozygotes, where the variant haemoglobin constitutes about 10–15% of the total haemoglobin present and Hb A most of the rest. Some homozygotes have however also been identified. Characteristically they have no Hb A or Hb A_2, and they generally show a moderate to severe degree of chronic haemolytic anaemia (Neeb et al. 1961, Duma et al. 1968).

The sequences of normal β and δ polypeptide chains differ from one another in 10 residues, at positions 9, 12, 22, 50, 86, 87, 116, 117, 125 and 126 (fig. 2.7, p. 36). Thus the nucleotide sequences in the corresponding genes are likely to be very similar over most of their lengths. Furthermore it is evident from the linkage studies discussed earlier that the two loci are on the same chromosome and probably lie very close to one another, if not in immediate

juxtaposition. Thus there exists a possibility that occasionally at meiosis the β-locus on one chromosome will pair aberrantly with the δ-locus on its partner. If following such a displaced synapsis crossing over were to occur in this region, then a new allele could arise which would specify a polypeptide chain with the δ sequence at one end and the β sequence at the other, and this is precisely the structure found in the Lepore haemoglobins.

Fig. 3.8 illustrates how such unequal crossing-over between the β- and δ-loci could give rise to two abnormal chromosomal products. In one chromosome the normal β- and δ-alleles would be absent, but a new gene

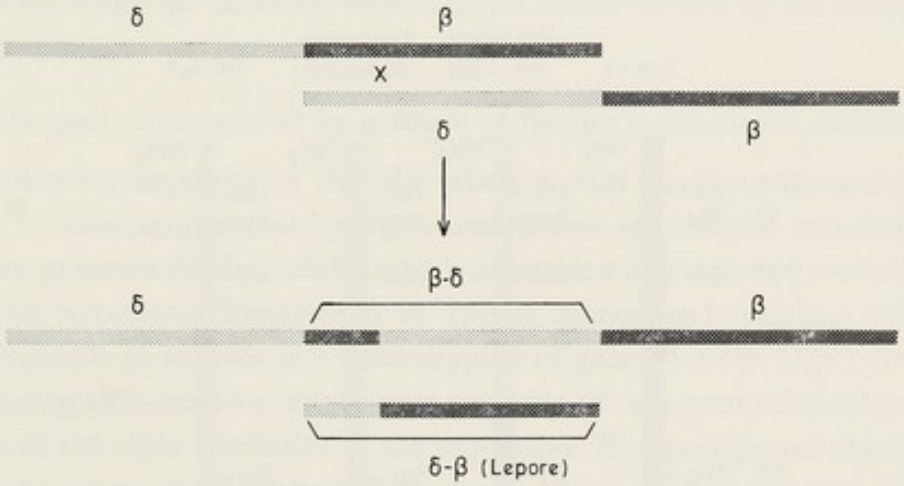


Fig. 3.8. Diagram illustrating how aberrant pairing of β and δ haemoglobin loci and crossing over may give rise to a 'Lepore' gene. The $\times$ indicates the site of crossing over.

(designated δ-β) would occur. In the other chromosome the normal β- and δ-alleles would be present and between them there would be a new gene (designated β-δ). Since both types of Lepore haemoglobins that have been studied in detail are associated with deficiencies of both normal Hb A ($\alpha_2\beta_2$) and normal Hb A_2 ($\alpha_2\delta_2$), it seems probable they are derived from the type of cross-over product in which the normal β- and δ-genes are lost. The other type of cross-over product would presumably be capable of causing the synthesis of both the normal β- and δ-chains as well as the abnormal β-δ chain and might therefore be very much more difficult to detect.

Structural studies on Hb-Lepore Hollandia indicate that the aminoacid residues up to and including number 22 in the sequence are the same as in the normal δ-chain, whereas residues from number 50 onwards are the same as in the normal β-chain (fig. 3.9). This indicates that the site of crossing-over must have been somewhere between the base triplets coding for these two

residues. The aminoacid sequence between position 22 and position 50 is the same in both the normal δ- and the normal β-chains. So crossing-over at many different sites in this region could result in identical products. This illustrates a general characteristic of unequal crossing-over, namely that it can yield the same products even with considerable variation in the actual position of the cross-over. In Hb-Lepore Washington the residues up to and including number 87 are identical with those in the normal δ-chain, and the residues from number 116 onwards are identical with those in the normal

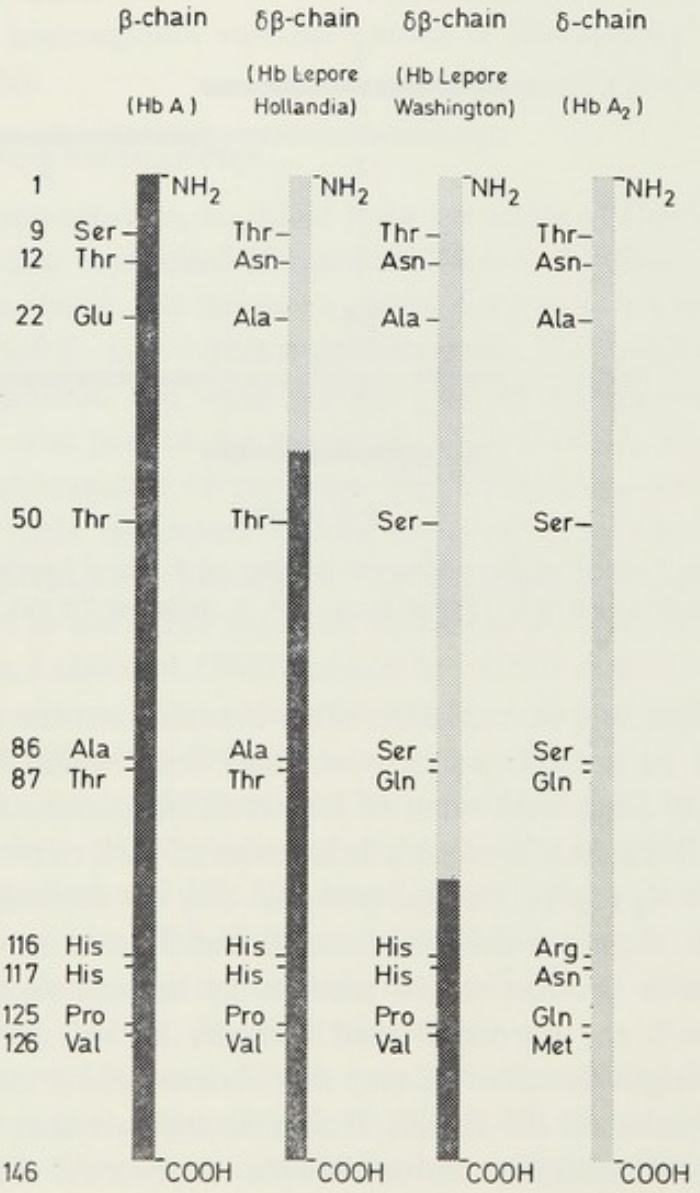


Fig. 3.9. Relationships of the aminoacid sequences of the non-α or $\delta\beta$-chains of Hb Lepore Washington and Hb Lepore Hollandia to those of the normal β and δ haemoglobin chains. Only the aminoacids which differ in the β and δ polypeptide sequences are shown.

β-chain. The sequence between positions 87 and 116 are the same in the two chains, and so in this case crossing-over at any point along this region could have produced the variant.

It will be noted that the results of structural studies on the Lepore haemoglobins strengthen the conclusion derived from family studies (see pp. 63–65) that the δ- and β-loci are closely linked and quite possibly lie immediately adjacent to one another on the same chromosome. They also provide some information about their probable order in the chromosome. This can be represented conventionally as $\delta\beta$ rather than $\beta\delta$, because the amino-terminal end of the Lepore haemoglobins is evidently determined by the first part of the δ-locus, while the carboxyl terminal end is determined by the second part of the β-locus.

3.3.2. Unequal crossing-over as a cause of further haptoglobin variants

The duplicated sequence of the Hp^2 allele would also be expected to predispose to misalignment at synapsis and hence to unequal crossing over. This may generate further alleles which determine haptoglobin α-chains with distinctive structures (Smithies et al. 1962b, Nance and Smithies 1963).

For example at meiosis in a heterozygote of genotype $Hp^{1F}Hp^2$, the Hp^{1F} allele in one chromosome might pair with the 'S' segment of the Hp^2 allele because of the close similarity of the sequences. If crossing-over should then happen to occur at an appropriate site in the paired chromatids then an Hp^{1S} allele could be formed as one cross-over product and a new version of the Hp^2 allele as the other (fig. 3.10). This new Hp^2 allele would be in effect an almost complete duplication of Hp^{1F}, and if we write the standard form of Hp^2 as $Hp^{2(FS)}$ then the new allele could be written $Hp^{2(FF)}$. In the same

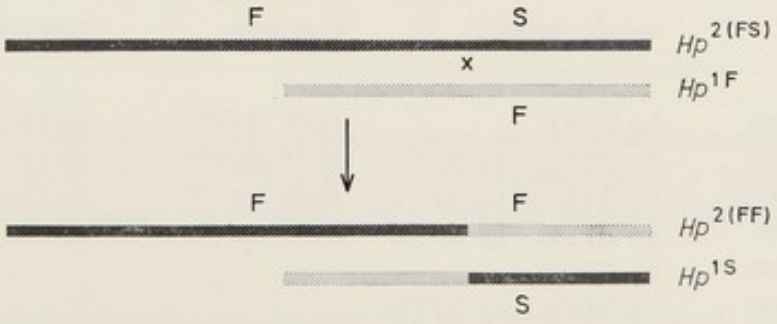


Fig. 3.10. Diagram illustrating how crossing-over in a heterozygote for the haptoglobin alleles $Hp^{2(FS)}$ and Hp^{1F} may result in the formation of the alleles $Hp^{2(FF)}$ and Hp^{1S}. F and S indicate the sites of the base differences in the DNA sequence of the haptoglobin genes which give rise to the lysine/glutamic acid substitutions in the polypeptides shown in fig. 3.3. The $\times$ indicates the site of crossing over.

way another possible form of the Hp^2 allele, namely $Hp^{2(SS)}$ could arise from crossing-over in a heterozygote $Hp^{2(FS)}Hp^{1S}$. Variants of the polypeptide hp2α with the electrophoretic properties expected of polypeptides with structures hpFFα and hpSSα have indeed been detected in certain individuals and it seems likely that they are the products of the alleles $Hp^{2(FF)}$ and $Hp^{2(SS)}$ which have arisen in just this way (Nance and Smithies 1963).

It will be seen that if this is so, four different versions of the Hp^2 allele may be expected to exist, and the hp2α polypeptide chains which they determine will differ according to whether lysine or glutamic acid is present at positions 54 or 113 in the aminoacid sequence. The four different alleles and the polypeptide chains they determine may be written as follows:

$Hp^{2(FS)}$ → hp 2(FS)α i.e. 54 Lys, 113 Glu
$Hp^{2(SF)}$ → hp 2(SF)α i.e. 54 Glu, 113 Lys
$Hp^{2(FF)}$ → hp 2(FF)α i.e. 54 Lys, 113 Lys
$Hp^{2(SS)}$ → hp 2(SS)α i.e. 54 Glu, 113 Glu

It follows that ten different genotypic combinations may be included in the so-called Hp 2-2 phenotype (table 3.1), although only some of these may be distinguishable electrophoretically even when the separated hp 2α chains are examined. Thus while hp 2(FF)α and hp 2(SS)α are probably distinguishable

TABLE 3.1

Possible Hp 2-2 and Hp 2-1 types.

Haptoglobin 2-2		Haptoglobin 2-1	
Genotype	α-polypeptides	Genotype	α-polypeptides
$Hp^{2(FS)}$ $Hp^{2(FS)}$	hp 2(FS)α	Hp^{1F} $Hp^{2(FS)}$	hp 1Fα+hp 2(FS)α
$Hp^{2(FS)}$ $Hp^{2(SF)}$	hp 2(FS)α+hp 2(SF)α	Hp^{1F} $Hp^{2(SF)}$	hp 1Fα+hp 2(SF)α
$Hp^{2(FS)}$ $Hp^{2(FF)}$	hp 2(FS)α+hp 2(FF)α	Hp^{1F} $Hp^{2(FF)}$	hp 1Fα+hp 2(FF)α
$Hp^{2(FS)}$ $Hp^{2(SS)}$	hp 2(FS)α+hp 2(SS)α	Hp^{1F} $Hp^{2(SS)}$	hp 1Fα+hp 2(SS)α
$Hp^{2(SF)}$ $Hp^{2(SF)}$	hp 2(SF)α		
$Hp^{2(SF)}$ $Hp^{2(FF)}$	hp 2(SF)α+hp 2(FF)α		
$Hp^{2(SF)}$ $Hp^{2(SS)}$	hp 2(SF)α+hp 2(SS)α	Hp^{1S} $Hp^{2(FS)}$	hp 1Sα+hp 2(FS)α
$Hp^{2(FF)}$ $Hp^{2(FF)}$	hp 2(FF)α	Hp^{1S} $Hp^{2(SF)}$	hp 1Sα+hp 2(SF)α
$Hp^{2(FF)}$ $Hp^{2(SS)}$	hp 2(FF)α+hp 2(SS)α	Hp^{1S} $Hp^{2(FF)}$	hp 1Sα+hp 2(FF)α
$Hp^{2(SS)}$ $Hp^{2(SS)}$	hp 2(SS)α	Hp^{1S} $Hp^{2(SS)}$	hp 1Sα+hp 2(SS)α

from one another and also from hp 2(FS)α and hp 2(SF)α with appropriate electrophoretic techniques, hp 2(FS)α and hp 2(SF)α are evidently indistinguishable (Nance and Smithies 1963). Similarly eight different genotypic combinations causing the Hp 2-1 phenotype may occur (table 3.1) though again not all of these may be distinguishable by electrophoresis. The differentiation of these various types is of course likely to be even more complex if as one would expect there also occurred occasionally other haptoglobin variants due to mutations of the single base change type and resulting in an aminoacid substitution in one or other of the α-chains with an alteration in charge. In general it seems that the alleles $Hp^{2(FF)}$ and $Hp^{2(SS)}$ are much less frequent than the other Hp^2 alleles ($Hp^{2(FS)}$ or $Hp^{2(SF)}$). Thus in one extensive population study (Nance 1967) it was found that although the total Hp^2 frequency was close to 0.50, the separate frequencies of $Hp^{2(SS)}$ and $Hp^{2(FF)}$ were each only be about 0.01.

In homozygotes Hp^2Hp^2, pairing at meiosis will presumably usually be of the ordinary type expected between any pair of alleles. However because of the close homology between the first and the second halves of the sequence of Hp^2, there is likely to be a significant tendency for occasional misalignment or displaced synapsis; that is the pairing of the first half of one allele with the second half of the other (fig. 3.11). If crossing over were then to occur

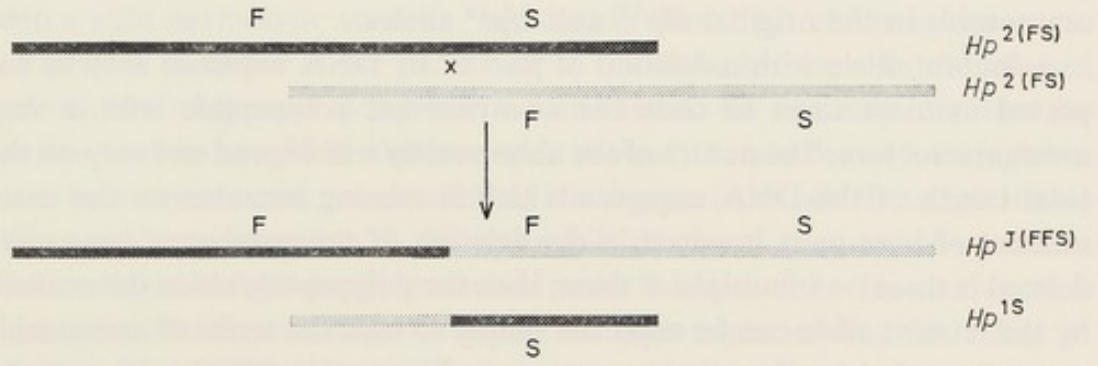


Fig. 3.11. Diagram illustrating how unequal crossing over between the haptoglobin alleles in a homozygote $Hp^{2(FS)}Hp^{2(FS)}$ may result in the appearance of a 'triplicated' allele $Hp^{J(FFS)}$ and a normal Hp^{1S} allele. The × indicates the site of crossing over.

in the paired segment, this could give rise to a new 'triplicated' allele, as well as an ordinary Hp^1 allele (Smithies 1964). Such a 'triplicated' allele would be expected to result in the formation of an hpα chain nearly three times the size of Hp1Fα or Hp1Sα. What appears to be a polypeptide chain of this size (Dixon 1966) has in fact been separated from a very unusual haptoglobin

type (the so-called 'Johnson' phenotype). Although the detailed structure of this extended polypeptide has not yet been determined, it seems quite likely that it represents the product of a 'triplicated' haptoglobin allele of the sort which may be expected to arise occasionally as a result of unequal crossing-over in the genotype Hp^2Hp^2.

3.4. Deletions

Certain types of mutational event can apparently result in the loss or deletion of part of a chromosome or chromatid, and the abnormality may then be transmitted to subsequent generations by the ordinary process of chromosome replication. Such mutations may be the consequence of two or more simultaneous chromosome or chromatid breaks with aberrant reunions, or they may arise because of unequal crossing over or from some other cause. Often they may result in the deletion of the DNA representing a whole series of genes, but in some cases the deletion may be much less extensive and involve only a sequence of base pairs within the confines of a single gene. For example the mutational event which originally gave rise to the Hp^2 allele (see pp. 72–75) at the α-haptoglobin locus, probably produced as its other product a partially deleted haptoglobin allele in which a sequence of 72 base pairs (24×3) occurred instead of the 249 (83×3) coding for the eighty-three aminoacids in the original Hp^{1F} and Hp^{1S} alleles.

A mutant allele with a deletion of part of its DNA sequence may be expected in most cases to code for an abnormal polypeptide with a very unusual structure. The nature of the abnormality will depend not only on the total length of the DNA sequence which is missing but also on the exact number of base pairs involved in the deletion. If the number of base pairs deleted is three or a multiple of three, then the polypeptide chain determined by the mutant allele can be expected simply to lack the series of aminoacids normally coded by the missing sequence of base pairs. That is to say, the sequence of aminoacids proximal to the deleted sequence will be the same as in the corresponding normal polypeptide chain and this will be immediately followed by the sequence of aminoacids which normally occur distal to the deleted segment. However if the number of base pairs deleted is not exactly three or a multiple of three, then an even more drastic alteration in the structure of the polypeptide for which it codes will be expected. This is because in the translation of the base pair sequence of the nucleic acid into the corresponding sequence of aminoacids in a polypeptide chain, the bases are read off three at a time consecutively, and if a number of bases which is not a

multiple of three are missing all the subsequent bases in the sequence will be misread and will in general now designate an entirely different aminoacid sequence. So such a mutation will be expected to result in an abnormal allele coding for polypeptide which only resembles its normal counterpart in the sequence of aminoacids proximal to the beginning of the deleted segment. Such mutations are said to involve 'frame shifts'.

Although a very large number of different alleles with partially deleted DNA sequences may be generated by different mutational events, the polypeptide structures that they define will in most cases be very unlikely to result in a viable protein which is capable of being detected. If the protein is formed at all, it will usually be extremely unstable and generally functionally inactive. Only in what must be a very small proportion of all the mutants of this sort that appear, can one expect to detect the abnormal protein product. Nevertheless certain haemoglobin variants have in fact been found in which the structural abnormality implies that they are determined by such partially deleted mutant alleles.

One example is Hb Freiburg, in which a single aminoacid residue has been shown to be missing (Jones et al. 1966). This is a valine normally present at position 23 in the β polypeptide chain. The rest of the aminoacid sequence of the β-chains, and also the complete sequence of the α-chain are apparently normal. The abnormal haemoglobin was discovered in a woman with a mild haemolytic anaemia, and also in two of her children, but it was not present in either of her parents or in her three brothers. In the affected individuals it occurred along with normal Hb A. So one may infer that the mother and those of her children who showed the abnormal haemoglobin were heterozygous for a rare allele at the β-gene locus which had probably arisen as a new mutation in the germ cells of one of the mother's parents. The mutation had presumably resulted in the deletion of the three consecutive base pairs which normally constitute the base triplet coding for valine at position 23 in the β polypeptide chain.

Another example is Hb Gun Hill (Bradley et al. 1967, Rieder and Bradley 1968) in which the abnormality also occurs in the β-chain. Here five aminoacid residues were found to be missing and although it could be concluded that they represented the sequence which normally occurs either in positions 91-95, or 92-96, or 93-97 in the β-chain, the exact position of the deletion could not be determined from the analytical data. The abnormal haemoglobin was found as well as normal Hb A in an adult with a long history of chronic haemolytic disease, and in his daughter. They were presumably heterozygous

for a rare allele at the β-locus, in which a sequence of fifteen consecutive base pairs had been lost.

Both Hb Freiburg and Hb Gun Hill belong to the class of 'unstable' haemoglobins (pp. 23–24), the instability presumably being a consequence of a distortion of the three dimensional configuration of the protein due to the missing aminoacids. The distortion of the three-dimensional structure was also indicated in other ways. Thus in the case of Hb Freiburg it was found that the spectral absorption curve of the methaemoglobin derivation was significantly different from that found with the met-forms of either Hb A or any of the Hb 'M's (p. 22) which have so far been identified. Although the site of the missing aminoacid (β 23) in the three-dimensional arrangement of the polypeptide chain is not apparently very close to the haem group, nevertheless the deletion evidently distorts the molecular structure in such a way as to modify the haem–globin interrelationships. It was also found that the heterozygous individuals with this abnormality showed a consistent though mild degree of cyanosis due to methaemoglobinaemia, presumably because the iron in the haems attached to the abnormal β-chains when oxidised to the ferric form is less readily reduced than in normal haemoglobin.

In the case of Hb Gun Hill, both the haems normally attached to the β polypeptide chains were found to be missing, so that the protein was functionally quite abnormal. The five aminoacids which are deleted in this haemoglobin normally occur in a region (β 90-97) closely associated with the haem group, and presumably their absence creates a situation in which a haem cannot be held in a stable association with the polypeptide chain.

3.5. Two separate aminoacid substitutions in a single polypeptide chain

Although mutations are very rare events, occasionally two separate mutations may come to be represented in a single allele. The two alterations will then be reflected in the structure of the polypeptide it determines.

An example of this is provided by the rare haemoglobin variant Hb Harlem (Bookchin et al. 1967) which was originally found along with normal Hb A in an otherwise healthy Negro. He was presumably heterozygous for an abnormal gene at the β-locus and for the normal β-allele, because his red cells, besides containing the variant haemoglobin, also contained Hb A, and because six of his nine children showed the same combination of haemoglobins. The β-chain of the variant haemoglobin has a substitution at position 6 of valine for glutamic acid, the same as occurs in sickle-cell haemoglobin;

and a substitution at position 73 of asparagine for aspartic acid which is the same as has been found by itself in another rare variant – Hb Korle-Bu (Konotey-Ahulu et al. 1968). The structure of Hb C Harlem may therefore be written: $\alpha_2\beta_2^{6\ Glu \rightarrow Val,\ 73\ Asp \rightarrow Asn}$. As in sickle-cell heterozygotes, red cells with the variant haemoglobin and Hb A showed the sickling phenomenon, and the variant like Hb S was found to be very insoluble in the deoxygenated state.

It is very unlikely that the allele coding for the unusual β polypeptide chain in Hb C Harlem originated from a single mutational event which led to two such well-separated single base alterations in the DNA sequence of the gene. Most probably it is the product of two quite distinct mutations occurring in different individuals living at quite different times. Since the sickle-cell allele is so relatively common in Negroes, the change at the β6 position was probably the earliest of the two mutations to occur. The other mutation is likely to have occurred in a much more recent generation.

It is of interest to note that the separate mutational alterations could have come to be present together in a single allele, in two somewhat different ways. The Hb C Harlem allele might have originated in an individual carrying the sickle-cell allele on one chromosome and the normal β allele on the other, by a mutation in the sickle-cell allele at the seventy-third base triplet. Alternatively it could have arisen in an individual who carried the allele determining Hb S ($\alpha_2\beta_2^{6\ Glu \rightarrow Val}$) on one chromosome and the allele determining Hb Korle-Bu ($\alpha_2\beta_2^{73\ Asp \rightarrow Asn}$) on the other, as a result of crossing-over somewhere between the sixth base triplet and the seventy-third base triplet during normal chromosome pairing at meiosis. Such a cross-over would have resulted in one of the recombination products coding for the abnormal β-chain of Hb C Harlem, and the other coding for the normal β-chain.

CHAPTER 4

Gene mutations affecting rates of protein synthesis

4.1. Genetic regulation of enzyme and protein synthesis

The genetical constitution of an individual not only defines the primary structures of all the many different enzymes and proteins which he is capable of synthesising, but it is also concerned with the regulation of their rates of synthesis. However the detailed manner in which such regulatory functions are exercised in higher organisms is in most respects still very obscure.

Some of the problems can be illustrated by considering the synthesis of haemoglobin in the normal individual. This protein is formed in quite considerable amounts by certain erythropoietic cells, but is not synthesised appreciably by the many other types of cells in the organism. Yet one supposes that all these cells contain the same complement of genes. So we must ask why in some cells the haemoglobin genes are effective and in others their activity is repressed. Even in cells which are capable of synthesising haemoglobin, quite marked disparities are observed. They all presumably contain the β-, γ- and δ-loci. Yet in the foetus the rate of synthesis of γ-chains is considerably greater than that of β-chains, while in the adult β-chain synthesis predominates and γ-chain synthesis is barely detectable. One also finds that δ-chain synthesis which occurs in the adult, goes on at about one fortieth the rate of β-chain synthesis. These relationships are illustrated in fig. 4.1.

The same kind of problem is posed by the distribution of enzyme proteins. Certain enzymes are found in most tissues of the organism and the particular proteins concerned appear to be the same in all cells. Others are much more localised. In some cases the specific enzyme activity appears to be confined to only one or a very limited number of tissues, and there are of course many examples where a particular enzyme though widely distributed appears to be formed at markedly different rates in different tissues. Furthermore what is apparently the same metabolic function may be performed by structurally distinct forms of an enzyme (isozymes) in different tissues. Liver phosphory-

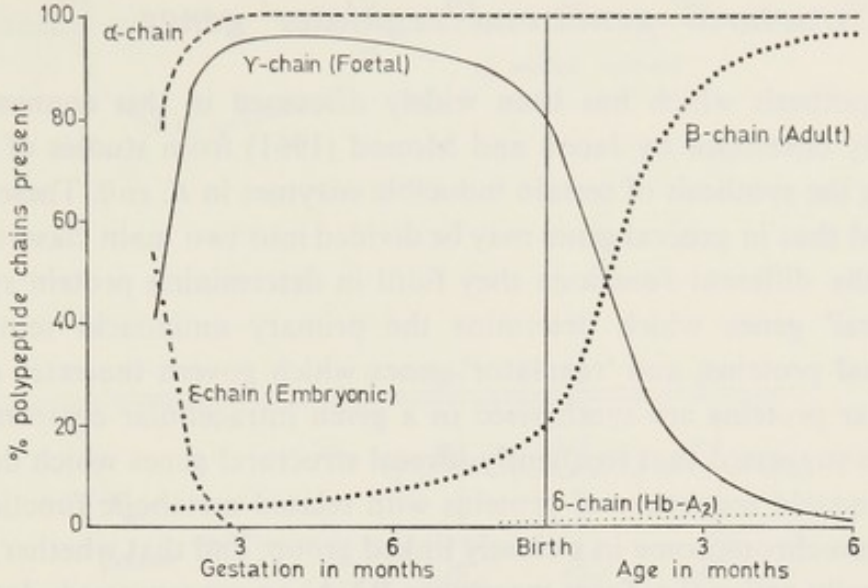


Fig. 4.1. The formation of the α, β, γ, δ and ε haemoglobin polypeptide chains in early life (Huehns et al. 1964).

lase for example differs in its structure from muscle phosphorylase and a gene mutation may affect one and not the other (see p. 164). Yet they each seem to be involved in the same way in glycogen degradation. Similarly enzymes such as aldolase and pyruvate kinase exist as structurally distinct proteins in different tissues, and a single mutation may affect one form and not another (see pp. 150–155). The variation of lactate dehydrogenase isozymes from one tissue to another (see pp. 40–41) illustrates the same general point.

All this is perhaps not very surprising because it is in effect simply another way of describing tissue differentiation as it occurs in multicellular organisms. Individual cells each exhibit a characteristic pattern of enzyme and protein synthesis, and they do not make the full range of proteins and enzymes of which the organism as a whole is capable. But since in general all nucleated cells in a single individual with the exception of the gametes are thought to have the same complement of genes, one must suppose that in any one cell the activity of a substantial fraction of these is continuously repressed.

Thus the problem of the genetic regulation of the rates of protein synthesis must be viewed as part of the wider problem of embryological development and of tissue differentiation. As yet no satisfactory general theory which adequately accounts for these phenomena has been advanced.

4.2. 'Structural' genes and 'regulator' genes

One hypothesis which has been widely discussed in this connexion was originally developed by Jacob and Monod (1961) from studies of mutants affecting the synthesis of certain inducible enzymes in *E. coli*. These authors proposed that in general genes may be divided into two main classes according to the different functions they fulfil in determining protein synthesis; 'structural' genes which determine the primary aminoacid sequence of individual proteins, and 'regulator' genes which govern the rates at which particular proteins are synthesised in a given intracellular environment. It was also suggested that frequently several structural genes which determine the aminoacid sequences of proteins with related metabolic functions may occur on a chromosome in a closely linked group, and that whether they are functionally active (i.e. form messenger RNA) or are repressed, depends on the state of a segment of DNA, called the 'operator', located at one end of the set of linked 'structural' genes. Such a group of adjacent 'structural' genes under the control of a single operator is referred to as an 'operon', and is thought to behave as a unit in the formation of messenger RNA (i.e. in transcription). The hypothesis suggests that the state of the operator and hence the activity of the associated structural genes is controlled by a repressor substance which is the product of a 'regulator' gene.

Originally there was some doubt about the chemical nature of the postulated repressor, but it was subsequently shown in more than one case that it is in fact a protein (Gilbert and Müller Hill 1966, Ptashne 1967). So one may presume that both 'regulator' and 'structural' genes code for the aminoacid sequences of specific proteins, and that the distinction between them essentially depends on the functional role of the proteins they determine.

The suggested interrelations between these different genetic units are illustrated by the model scheme shown in fig. 4.2. S_1, S_2 and S_3 are structural loci whose DNA base sequences define the aminoacid sequences in the polypeptide chains P_1, P_2 and P_3. M_1, M_2 and M_3 are the intermediary messenger RNA's, probably formed as a single continuous strand. O is the operator and lies immediately adjacent to S_1. O, S_1, S_2 and S_3 together constitute an 'operon'. RG is a regular gene which forms the repressor protein R. R is capable of binding specifically with the operator O and when this occurs structural gene activity along the whole length of the operon is repressed. R however is also capable of complexing with certain small molecules (inducers) which may be present in the intracellular milieu. When this happens its configuration is altered in such a way that it no longer acts on the

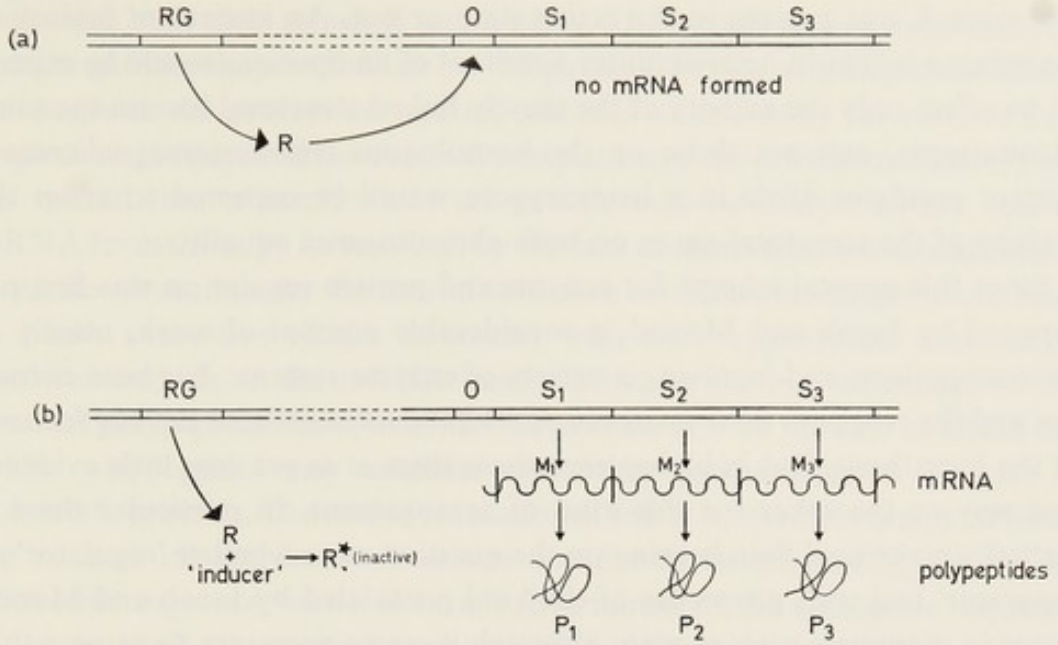


Fig. 4.2. Diagram illustrating the regulation of 'structural' gene activity according to the scheme of Jacob and Monod (1961), as described in the text. (a) Repression of 'structural' gene activity. (b) Derepression of 'structural' gene activity.

operator, thus allowing structural gene activity to proceed. So the presence of the particular metabolites which complex in a specific way with R, will affect the synthesis of all the polypeptides (P_1, P_2 and P_3) coded by the 'structural' genes in the operon. If such a metabolite is absent the structural genes will be repressed. If it is present they will show more or less activity depending on its concentration and its affinity for R.

Mutation at one of the structural loci (e.g. S_2) will usually affect only the synthesis of the corresponding polypeptide chain (P_2), altering its structure by causing, for example, a single aminoacid substitution. Mutation at the regulator locus on the other hand will be expected to affect the rates of synthesis of all the polypeptides determined by the structural genes in the operon, but will not alter their structures. Thus it might result in a modification of the repressor substance R such that it can no longer act on the operator. If so there would be continuous and uncontrolled synthesis of P_1, P_2 and P_3 irrespective of changes in the intracellular milieu. Alternatively it might modify R so that it no longer complexes with the inducer molecules but can still act on the operator. If so continuous repression of synthesis of the polypeptides would occur. Mutations of the operator may also be envisaged. According to their nature they might result in continuous repression or continuous activation of the structural genes, irrespective of whether the

repressor R was present in the active state or not. An important feature of the scheme is that in heterozygotes a mutant of an operator would be expected to affect only the activity of the closely linked structural loci on the same chromosome, and not those on the homologous chromosome; whereas a mutant regulator allele in a heterozygote would be expected to affect the activity of the structural genes on both chromosomes equally.

Since this general scheme for enzyme and protein regulation was first put forward by Jacob and Monod, a considerable amount of work, mainly in microorganisms and involving a variety of enzyme systems, has been carried out and the results in their main essentials have substantiated the key features of the hypothesis. But in higher organisms there is as yet very little evidence one way or the other for this kind of arrangement. In particular there is virtually no critical data bearing on the question as to whether 'regulator' or 'operator' loci with properties of the kind postulated by Jacob and Monod occur in mammals such as man, although it seems necessary to assume that some sort of rather precise genetical control of the rates of synthesis of enzymes and proteins must exist.

A considerable variety of inherited disorders involving specific deficiencies of particular enzymes and proteins are known in man, and it has been suggested at various times that a number of these may be due to mutations of regulator genes or of operators. In some instances where this has been postulated, subsequent work has shown that in fact a structural abnormality of the particular protein occurred and so the mutation must have affected the so called 'structural' gene for the proteins (e.g. in acatalasia – Aebi et al. 1968). In the other cases the available data is generally inadequate to enable a critical distinction between the different possibilities to be made.

4.3. Structure-rate relationships

Another possibility which must be considered in connexion with the general problem of the genetic control of rates of synthesis of specific proteins is that the nucleotide sequence of a given 'structural' gene besides coding for the aminoacid sequence of a polypeptide chain may also in some degree determine the rate at which it is synthesised (Itano 1957, 1965, 1966; Boyer et al. 1964).

In the synthesis of a specific polypeptide chain the DNA chain coding for its aminoacid sequence is first transcribed into a corresponding mRNA chain which then becomes attached to ribosomes where it serves as a template for the formation of the polypeptide chain by the sequential binding of tRNA

molecules with their attached aminoacids. The overall rate of synthesis of a specific protein may be limited by the particular rates at which these different stages in the process take place, and also of course by the rates of subsequent stages required for the final completion of the protein molecule. Thus the overall rate of synthesis of the protein may be limited by the rate of mRNA formation, by its stability, by the rate of its attachment to and release from the ribosomes, by the rate of initiation of tRNA binding and the rates at which successive tRNA molecules with their appropriate aminoacids bind successively to the mRNA base triplets, by the rate of release of the finished polypeptide chain from the ribosomes, and by the rate at which the folded polypeptides may combine with other polypeptides in the case of polymeric proteins. The rate of formation of mRNA, its stability and its binding to and release from ribosomes could well be affected by its precise nucleotide sequence, which in turn is determined by the nucleotide sequence of the particular 'structural' gene concerned. Furthermore the rate of polypeptide chain formation on the ribosomes could be affected by the particular mRNA sequence, or the availability of the different tRNA molecules needed to assemble the polypeptide chain. Also the particular sequential arrangement of the aminoacids in the protein could affect the rate at which it folds into an appropriate configuration as the chain is assembled, and also perhaps the rate at which it is released from the ribosomes. Finally the precise polypeptide structure may affect the rate at which it is capable of combining with other polypeptides or prosthetic groups to give the final protein. Thus the specific base sequences of the DNA of any particular 'structural' gene or genes could in a variety of ways and at several different levels influence the rate of synthesis of a particular protein.

Heterozygotes for 'structural' gene mutations offer one approach to investigating whether this kind of effect actually occurs. It has been suggested for example that differences in the amounts of the two polypeptide products observed in certain heterozygotes, may in some cases be due to differences in their rates of synthesis imposed by the difference in base sequences of the two alleles. Heterozygotes with the sickle-cell trait and also those with the haemoglobin C trait illustrate the general argument. In the sickle-cell trait it is consistently found that there is more Hb A in the red cells than Hb S. On average about 60–65% of the haemoglobin present is Hb A ($\alpha_2\beta_2$) and about 35%–40% is Hb S ($\alpha_2\beta_2^{6\ Glu\rightarrow Val}$) (Wells and Itano 1951, Wrightstone and Huisman 1968). A similar disproportion is observed in haemoglobin C trait, where again the amount of Hb A present is significantly more than the amount of Hb C. The total amount of haemoglobin formed in these hetero-

zygous states is not on average significantly different from that in normal homozygotes. Furthermore although there may be minor differences in stability between the variant haemoglobins and Hb A, they would appear to be too slight to account for in terms of preferential breakdown of the variant haemoglobin for the disparity observed. It seems then that the characteristic disproportion in the amounts of the two haemoglobins in these heterozygotes is quite possibly a consequence of differences in rates of synthesis (Itano 1965, Boyer et al. 1964). This implies that the rate of synthesis of for example the $\beta^{6\ Glu \rightarrow Val}$ polypeptide chain in the sickle-cell trait, or the $\beta^{6\ Glu \rightarrow Lys}$ polypeptide chain in the haemoglobin C trait, is significantly less than the rate of synthesis of normal β polypeptide chains in these heterozygotes. The mutation in each case is thought to alter by a single base change the nucleotide sequence of the messenger RNA, and to alter by a single aminoacid substitution the aminoacid sequence of the corresponding polypeptide. Either of these changes could in theory limit the overall rate of synthesis, as compared with that of the normal polypeptide. But the exact nature of the proposed structure–rate relationship in these cases remains obscure.

Such structure–rate relationships could well be relevant not only to the synthesis of variant polypeptides determined by mutant or abnormal genes, but also to so-called normal alleles. Indeed it is possible that in any gene there may be one or just a few triplets in the nucleotide sequence which by the restrictions they impose at some step in the process of polypeptide synthesis, are rate limiting. So they will effectively determine the maximal overall rate of synthesis of the particular polypeptide. Relevant to this is the fact that any one aminoacid may be coded by several distinct mRNA triplets which may require different tRNA molecules. So the triplet which actually codes for a particular aminoacid in a sequence, and the availability of the corresponding tRNA in the cell, could be important in deciding the overall rate of synthesis of the polypeptide. An interesting case where marked disparities in the rates of assembly of two distinct polypeptide chains determined by separate loci, have been demonstrated directly is provided by the normal β- and δ-chains of haemoglobin (Winslow and Ingram 1966). These chains differ in only ten aminoacids so the genes coding for them must be very similar. Yet under normal circumstances the rate of assembly of the β-chain is much greater than that of the δ-chain, though it is not known exactly how this comes about.

But whatever the detailed mechanisms involved, the idea that a gene by the very nature of its nucleotide sequence may provide an inbuilt control on the rate of synthesis of the polypeptide it determines may well prove an important one. It could in principle account for many of the characteristic differen-

ces that are observed between the amounts of different enzymes or proteins in a single cell.

4.4. Inherited defects in rates of protein synthesis - the thalassaemias

Quite a number of inherited abnormalities are known in which the central defect is the deficiency of a particular protein, and many of these are probably due to mutations which result in a gross but specific reduction in the rate of synthesis of one or more polypeptide chains. The most extensively studied of such conditions are those which involve defects in the synthesis of haemoglobin. Among them are a series of chronic haemolytic anaemias collectively known as the thalassaemias. They appear to be determined by a series of distinct abnormal genes, in different heterozygous and homozygous combinations. But it has not yet proved possible to define in any one case the precise nature of the mutation responsible, or the manner in which this results in the observed defect in haemoglobin synthesis. However since the conditions are probably representative of a whole class of similar disorders in which specific defects in rate of synthesis of a variety of other proteins and enzymes occur, the problems and difficulties posed by the thalassaemias are of some general relevance.

It is useful to classify the various thalassaemias according to the polypeptide chain primarily concerned in causing the haemoglobin deficit (Ingram and Stretton 1959). Thus in the so-called β-thalassaemias, the main defect appears to involve β-chain synthesis, whereas in α-thalassaemias α-chain synthesis is primarily affected.

Detailed reviews of the various thalassaemic conditions have been given by Weatherall (1965, 1969) and Motulsky (1964).

4.4.1. β-Thalassaemia

β-Thalassaemia occurs relatively commonly in certain populations in the Mediterranean countries (e.g. in Southern Italy and Greece) and it is also not uncommon among populations living in India and the Far East. It is probable that a number of different mutant genes can cause this type of abnormality and although they each result in depression of β-chain synthesis, the degree to which this occurs appears to vary considerably from one gene to another. In some cases β-chain synthesis may be completely or almost completely absent. In others synthesis of β-chains does occur but at a much reduced rate.

Homozygotes for such genes show a severe form of anaemia classically known as thalassaemia major (or Cooley's anaemia). The disease generally becomes manifest shortly after birth, at the time when γ-chain synthesis is normally giving place to β-chain synthesis. The severe anaemia results from a deficiency of haemoglobin A ($\alpha_2\beta_2$). This produces a stress situation in the erythroid marrow which by a mechanism not yet understood, leads to the continued production of cells still capable of forming haemoglobin F ($\alpha_2\gamma_2$). Consequently most of the haemoglobin present is Hb F. The remainder is Hb A (or very closely resembles it) but in the most severe cases this may be completely or almost completely absent. The continued production of γ-chains is generally quite insufficient to compensate for the deficiency of β-chain synthesis. So the mean red cell haemoglobin concentration is markedly reduced. Also the total number of red cells in the circulation is much less than normal, and the red cells themselves show very marked variations in shape and size.

Studies on the kinetics of globin synthesis in red cells in β-thalassaemia (Weatherall et al. 1965, Bank and Marks 1966, Huehns and Modell 1967, Bargellesi et al. 1967) have demonstrated the grossly reduced rate of β-chain formation and have also shown that α-chain production greatly exceeds β- and γ-chain formation. Thus a great excess of α-chains is produced. These are evidently very unstable and tend to precipitate and become associated with the red cell stroma. This in turn renders the red cells particularly susceptible to premature destruction. Consequently the severe anaemia characteristic of β-thalassaemia results not only from an overall deficit in haemoglobin synthesis, but also from an excessive rate of red cell destruction.

Heterozygotes often show a mild anaemia (thalassaemia minor), which is however very variable in degree and indeed may often not be clinically apparent. There are usually typical abnormalities of red cell morphology (microcytosis, anisocytosis, and target cells). Most of the haemoglobin present is haemoglobin A, but characteristically the proportion of haemoglobin A_2 ($\alpha_2\delta_2$) is increased, assuming values of perhaps 4 to 7% of the total instead of the normal 2 to 3% (Kunkel et al. 1957). Thus δ-chain synthesis appears to be somewhat augmented. Hb F may also occur in slightly increased amounts (0.5–4% of the total haemoglobin), and it is of interest that it is unevenly distributed between different red cells.

Heterozygotes for a β-thalassaemia gene may be heterozygous as well for one or another of the genes which determine variant haemoglobins with abnormal β-chains such as Hb S, Hb C or Hb E. The anaemias that result are referred to as thalassaemia–sickle-cell disease, thalassaemia–haemoglobin

C disease etc. In these conditions the variant haemoglobin predominates and accounts for 70% or more of the total haemoglobin present. It appears that in such cases synthesis of the variant β-chain (i.e. β^S, β^C or β^E) is not depressed, so that the proportion of Hb A found gives some indication of the degree of depression of synthesis of the normal β-chain caused by the particular β-thalassaemia gene present. In some cases normal β-chains may be virtually absent.

Two types of hypotheses have been advanced about the nature of the mutations that result in the defective β-chain synthesis characteristic of β-thalassaemia. One is that the mutations occur at the structural locus which determines the aminoacid sequence of the β-chain. In this case the β-thalassaemia genes would be allelic to the genes which determine β-chain variants such as Hb S or Hb C. The other hypothesis is that the mutations affect a different locus, which however is specifically involved with the regulation of the rate of β-chain synthesis in the normal individual.

To some extent information about this question can be obtained from studies of families in which children of doubly heterozygous individuals (e.g. individuals with either thalassaemia–sickle-cell disease or thalassaemia–haemoglobin C disease) occur. A considerable number of such informative families (Motulsky 1964, Weatherall 1965) have now been investigated and it is clear that with only a few possible exceptions the great majority of the children have received from their doubly heterozygous parent either the β-thalassaemia gene or the gene for the β-chain variant, but not both or neither (table 4.1). The pedigrees therefore suggest that the genes are either at the same locus (i.e. are allelic), or if they do occur at separate loci then these must be relatively closely linked, if not immediately adjacent to one another on the same chromosome. Since the doubly heterozygous parent in these different families appeared to transmit one or other of the abnormal genes but not both to any particular child, then we must conclude that if two loci are indeed involved then the doubly heterozygous parent in every case carried the two abnormal genes on the separate members of the particular pair of homologous chromosomes (i.e. in repulsion).

But it is not possible from the available data to distinguish critically between allelism and the alternative possibility that two separate but closely linked loci are concerned. It is important to note that studies of this kind are beset with considerable difficulties. If linkage is close a very large number of informative families need to be studied to identify a cross-over, and confusion because of undetected illegitimacy can easily occur. There are also special problems because of the variability in manifestation of the β-thalassaemia

TABLE 4.1

Offspring of matings between normal individuals and individuals heterozygous for both β-thalassaemia and a β-haemoglobin variant. (Further data from other workers are tabulated in Motulsky, 1964 and in Weatherall, 1965.)

Mating	No. of matings	Offspring				Total	References
		Presumptive cross-overs		Presumptive non-cross-overs			
		Abnormal β Hb-β thal.	Normal	Heterozygous for β-thal. only	Heterozygous for abnormal Hb only		
Hb S-β thal. × normal	7	—	? 1	13	12	26	Ceppellini (1959)
Hb S-β thal. × normal	6	—	—	7	11	18	Weatherall (1965)
Hb C-β thal. × normal	2	—	—	6	4	10	Weatherall (1965)
Totals	15	—	? 1	26	27	54	

genes, and the absence of any unequivocal criterion of their presence or absence. Thus it may prove difficult to distinguish with certainty between a milder manifesting thalassaemia heterozygote and a normal homozygote. It could also prove difficult, and in certain cases perhaps impossible, to distinguish phenotypically between an individual who is a simple heterozygote for a β-thalassaemia gene which results in severe or perhaps complete suppression of β-chain synthesis, and one who is doubly heterozygous and carries this gene and also a gene determining a β-chain variant on the same chromosome (i.e. in coupling). Such a doubly heterozygous combination might be expected to result in a complete or nearly complete suppression synthesis of the β-chain variant and so the individual could resemble an ordinary β-thalassaemia heterozygote (Motulsky 1964). These situations are of course those precisely involved in the detection of presumptive cross-overs.

If the β-thalassaemia gene does indeed occur at a locus separate from the β-chain structural locus but closely linked to it, then the findings in double heterozygotes such as β-thalassaemia–sickle-cell disease or β-thalassaemia–haemoglobin C disease indicate that the thalassaemia gene can only repress activity of the β-structural locus on its own chromosome. This is so because in these conditions the rate of synthesis of the β-chain variant (e.g. Hb S or Hb C) is apparently not affected. Thus on this hypothesis the β-thalassaemia locus might correspond in the Jacob–Monod scheme to an operator which was specifically concerned with controlling the activity of the adjacent β-chain structural gene. It may be noted that since δ-chain synthesis is not repressed in β-thalassaemia it would be necessary to suppose that the closely linked β- and δ-structural loci are not under the control of the same operator, and thus do not constitute an operon.

If, on the other hand, the β-thalassaemia genes represent mutations at the locus which determines the structure of the β-chain, the question arises as to exactly how such mutants result in the marked depression in β polypeptide synthesis which is observed. In theory a change in the nucleotide sequence of the β 'structural' gene might produce such an effect by imposing a severe rate limiting restriction at one of the several different stages in the normal processes of transcription and translation which lead to the final synthesis of the polypeptide and its incorporation into haemoglobin (see discussion of structure–rate relationships pp. 94–96). But the identification of such abnormalities poses exceedingly difficult technical problems which still in large part remain to be resolved, although some very suggestive results have been obtained (Clegg et al. 1968, Braverman and Bank 1969). Furthermore different β-gene mutants could affect the synthetic process in different ways and yet

produce a similar end result, that is a severe limitation in the rate of β-chain synthesis or its complete failure.

4.4.2. Hereditary persistence of foetal haemoglobin

It is of interest to contrast the effects of the β-thalassaemia genes with those of another mutant which results in complete suppression of β-chain synthesis. This mutant is usually referred to as the 'high Hb F' gene and it gives rise to a condition known as 'hereditary persistence of foetal haemoglobin ($\alpha_2\gamma_2$)'. It appears from family studies that the gene occurs at a locus which, like the β-thalassaemia locus, is either closely linked to the β-chain 'structural' locus or is in fact identical with it. It has mainly been observed in Africans or in people of African descent.

In adults heterozygous for this 'high F' gene and its normal allele (Conley et al. 1963), about 25% of the total haemoglobin is Hb F ($\alpha_2\gamma_2$). The rest is mainly Hb A ($\alpha_2\beta_2$) and the total amount of haemoglobin present is within the normal range. Some Hb A_2 ($\alpha_2\delta_2$) also occurs, but the proportion is somewhat less than is normally found. Thus δ-chain synthesis appears to be depressed, a situation which is different from that in heterozygotes for β-thalassaemia genes where the proportion of Hb A_2 is usually increased. Another difference is that the foetal haemoglobin apart from being much greater in amount than in β-thalassaemia heterozygotes is distributed evenly among the red cells, whereas in β-thalassaemia heterozygotes it is very unevenly distributed.

The homozygote for this 'high F' gene is evidently only capable of forming Hb F (Wheeler and Krevans 1961, Baglioni 1963). No Hb A or Hb A_2 are present, but compensation by Hb F synthesis is evidently quite adequate so that severe anaemia does not result.

Individuals heterozygous for the 'high F' gene and heterozygous as well for one of the genes determining a β-chain variant such as β^S or β^C have also been identified (Edington and Lehmann 1955, MacIver et al. 1961). Here the only haemoglobins present are the variant haemoglobin, and Hb F. There is no Hb A or Hb A_2.

Thus the abnormality produced by this 'high F' gene appears to represent a complete suppression of both β- and δ-chain synthesis, associated with a continuing synthesis of γ-chains which more or less completely compensates for the deficit of β and δ. This contrasts strikingly with the situation with the β-thalassaemia genes where usually only β-chain synthesis is suppressed and not δ-chain synthesis, and where compensatory synthesis of γ-chains although it occurs is very much less efficient so that severe anaemia results.

The nature of the mutational change which has given rise to the so called 'high Hb F' gene is not known. One hypothesis is that it represents a mutation of an 'operator' which controls both the β- and δ-chain structural genes, and which also may be somehow involved in the switch process which in normal individuals allows γ-chain synthesis to give way to β-chain synthesis during development. If this is so then the β and δ structural loci would be part of a single operon, a conclusion which would tend to make less plausible the hypothesis that β-thalassaemia is due to a mutation of an operator controlling β-chain synthesis but not δ-chain synthesis. Another hypothesis is that the so called 'high Hb F' gene actually represents a chromosomal deletion or rearrangement which affects both the β and the δ structural loci and effectively eliminates their activity.

4.4.3. Other abnormal genes affecting β- and δ-chain synthesis

A number of other apparently specific hereditary abnormalities involving defective synthesis of the β- and δ-haemoglobin chains have also been identified. One of these is known as $\beta\delta$- or $\delta\beta$-thalassaemia (Brancati and Baglioni 1966, Comings and Motulsky 1966) and has also been referred to as F-thalassaemia (Motulsky 1964, Weatherall 1965). It resembles β-thalassaemia clinically but tends to be less severe. However, both β- and δ-chain synthesis is depressed. Compensatory γ-chain synthesis occurs and is significantly more effective than in β-thalassaemia, though it does not completely compensate for the deficit of β-chains.

A further distinct condition in which suppression of both β- and δ-chain formation is found has been regarded as another and somewhat less marked form of 'hereditary persistence of foetal haemoglobin' (Fessas and Stamatoyannopoulos 1964). It is not uncommon in Greece and has been referred to as the 'Greek type', in contradistinction to the more marked form known as the 'African type'. β- and δ-chain synthesis is not completely abolished, but goes on at a much reduced rate. Increased γ-chain synthesis occurs and compensates more or less completely for the β-chain deficit. An important difference between this abnormality and the $\beta\delta$-thalassaemia referred to above is that in heterozygotes, although Hb F may comprise on average much the same proportion of the total haemoglobin present, it is evenly distributed between the circulating red cells, whereas in $\beta\delta$-thalassaemia Hb F is very unevenly distributed between different red cells (table 4.2).

Various hypotheses similar to those advanced to explain β-thalassaemia and the so-called 'African type' of 'hereditary persistence of foetal haemoglobin' have been put forward to account for these other different though

TABLE 4.2

Effects of different mutations of the Hb δ-β complex on the formation of HbF ($\alpha_2\gamma_2$) in heterozygotes (Motulsky 1964).

Type of mutation		Effect of mutation on		HbF ($\alpha_2\gamma_2$) in heterozygotes		
		β-chain synthesis	δ-chain synthesis	% of total Hb	Distribution of HbF among erythrocytes	Mean corpuscular haemoglobin in heterozygotes
'Thalassaemia'	β-thalassaemia	Mild to severe reduction	Increased	Small increase	Uneven	Low
	$\delta\beta$-thalassaemia	Mild to severe reduction	Normal or slight reduction	3.5–36%	Uneven	Low
'Hereditary persistence of foetal haemoglobin'	'Greek' type	Moderate reduction	Moderate reduction	8–16%	Even	Normal
	'African' type	Complete suppression	Complete suppression	16–36%	Even	Normal

apparently related conditions. But at present there is no critical way of deciding between the various theoretical possibilities. It is clear however that there must occur a variety of distinct mutants affecting in different ways the synthesis of β-, δ- and γ-chains, and giving rise to these different abnormalities. The precise way in which they produce their effects, however, remains obscure. One may anticipate that similar complexity and heterogeneity will be encountered in connexion with defects in rates of synthesis of other proteins and enzymes.

4.4.4. α-Thalassaemia

Since α-chains occur in both adult and foetal haemoglobins a mutation causing severe depression in α-chain synthesis would be expected to become manifest in foetal life. In fact a condition of this sort has been recognised. The affected foetus is grossly oedematous, a condition known as hydrops foetalis, and there is enlargement of the liver and other abnormalities (Lie-Injo et al. 1962, Pootrakul et al. 1967a). Death occurs *in utero* with abortion in the latter part of the pregnancy. Nearly all the haemoglobin present in the red cells has an abnormal structure. It appears to be a tetramer made up entirely of normal γ-chains and its structure may be written as γ_4 (Hunt and Lehmann 1959). It is usually referred to as Hb Bart's (Ager and Lehmann 1958). Evidently there is an almost complete absence of α-chain synthesis, but γ-chain synthesis proceeds normally and in the absence of α-chains the γ-tetramer is formed. The severe pathological consequences are due both to the haemoglobin deficit and also to the abnormal oxygen dissociation curve of Hb Bart's (γ_4) which tends to make oxygen less readily available to the tissues.

The condition has been attributed to homozygosity for a mutant gene either at the locus coding for the α-polypeptide chain, or at a locus which in some way controls its synthesis. Individuals heterozygous for this gene and its normal allele are generally healthy, and have at most only a minimal anaemia (Pootrakul et al. 1967a). In the newborn heterozygote small amounts of Hb γ_4 are present (5–10% of the total haemoglobin) but this tends to disappear along with Hb F ($\alpha_2\gamma_2$) during the following two months.

Another condition, known as Hb H disease, is also caused by defective α-chain synthesis (Motulsky 1964, Weatherall 1965). The disease is a variable though often moderately severe anaemia in which 10–30% of the haemoglobin present is an unusual form of haemoglobin known as Hb H, while the remainder is Hb A. Hb H is a tetramer made up of normal β-chains, i.e. β_4 (Hunt and Lehmann 1959). It is very unstable and gives rise to intraerythrocyte

inclusion bodies which are a characteristic feature of the anaemia. In a few cases it has been possible to study the cord blood of infants who subsequently developed Hb H disease (Pootrakul et al. 1967b). About 25% of the haemoglobin was found to be Hb γ_4, and this subsequently decreased progressively with a reciprocal increase in Hb β_4. Thus in Hb H disease there appears to be a partial deficiency of α-chains, which results in a relative excess of γ-chains in foetal life and of β-chains later on. Family studies (Pootrakul et al. 1967a, b) suggest that individuals with Hb H disease are heterozygous for the α-thalassaemia gene which in homozygotes results in hydrops foetalis, and also heterozygous for some other gene (possibly an allele) which depresses α-chain synthesis but not to the same degree.

CHAPTER 5

Quantitative and qualitative variation of enzymes

5.1. Quantitative variation in enzyme activity

Many gene mutations are expressed by a characteristic alteration in the activity of a specific enzyme. In some cases individuals homozygous for the mutant gene appear to have a complete or nearly complete deficiency of the particular enzyme activity. In other cases the reduction in activity produced by the mutant is less profound, and very occasionally increased activity has been noted. In fact it appears that almost any degree of alteration of an enzyme activity may be produced by different mutations.

The assessment of such changes in different individuals is in general based on quantitative measurements of the level of the enzyme activity in extracts of cells from a particular tissue or tissues (e.g. red cells, white cells, liver etc.). In considering the significance of these observations it is important to recognise that the measured level of activity of an enzyme in the cells of a tissue is a complex parameter. In the first place it depends on the specific catalytic properties of the enzyme, and these are determined by the detailed molecular structure of the enzyme protein. Secondly, it depends on the actual quantity of enzyme protein present in the tissue. This represents the resultant of the rates of two quite distinct and opposite processes; the rate of synthesis of the protein, and the rate at which it is broken down or decays. Finally extraneous factors such as the presence of specific inhibitors, cofactors, metal ions and so on may affect the observed activity, as well of course as the particular experimental conditions used in the assay procedure (e.g. substrate concentrations, *p*H, etc.). Other complicating factors are the localisation or compartmentalisation of the enzyme within the cell, and the particular methods used to disintegrate the cell for enzyme analysis.

It is apparent, then, that a particular mutant gene might cause an observed alteration in the level of enzyme activity in a variety of different ways. The main possibilities can be conveniently summarised as follows:

(a) The mutational change may have led to the synthesis of a structurally

altered protein with defective or modified catalytic properties. A single aminoacid substitution, for example, if it affected the active centre of the enzyme might, by altering the facility with which substrate or coenzyme is bound to the protein during catalysis, result in a marked modification in the kinetics of the catalytic process, and hence in the apparent enzyme activity. In such circumstances the actual quantity of enzyme protein present could be unaffected, even though the enzyme activity is reduced or perhaps completely absent.

(b) The mutant gene may lead to the synthesis of a structurally altered protein whose catalytic activity is not significantly affected, but whose inherent stability is less than that of its normal counterpart. The enzyme protein would tend to be more rapidly denatured *in vivo*, and its half life consequently reduced. The increased rate of breakdown would mean that the actual quantity of functionally active enzyme protein present at any one time, and hence the level of activity, would be correspondingly diminished. It might indeed be so low as to be undetectable.

(c) The rate of synthesis of the enzyme protein may be altered. This could occur following mutations in the so-called 'structural' gene (or genes) coding the aminoacid sequences of the protein, or it could result from a mutation in another gene, normally concerned in regulating the activity of such a 'structural' gene. The enzyme protein may simply not be formed, or may be formed at a reduced rate so that the quantity present at any given time is less than normal. Occasionally an increased rate of synthesis might result, and lead to an apparent enhancement of the enzyme activity.

(d) The particular enzyme may be affected only indirectly. A mutation occurring in a gene not normally involved in controlling the synthesis of the enzyme itself, may nevertheless influence its activity by for example causing an alteration in the intracellular concentration of some activator or inhibitor.

Some of these possibilities are of course not mutually exclusive and so in certain cases a mutant might bring about an altered enzyme activity in more than one way. For example, a single aminoacid substitution or other kind of structural change in an enzyme protein could lead to an alteration in its kinetic properties and also in its inherent stability.

These different possibilities, as well as a variety of technical problems make the analysis of the genetical and biochemical nature of inherited quantitative variation in enzyme activity peculiarly difficult. The technical problem largely arises because enzyme proteins, unlike a protein such as haemoglobin, generally occur only in trace amounts, so they cannot easily be

isolated and characterised from single individuals. Also they are usually only readily detected and specified by virtue of their catalytic activity, and if this is considerably reduced or even absent the elucidation of the underlying abnormality is not easy to investigate.

There is a further general problem which commonly arises in the study of situations where inherited differences between individuals in the level of activity of a particular enzyme have been identified. Although occasionally it is possible to classify people into sharply distinct groups in terms of enzyme activity, for example those with relatively high and those with relatively low activity (or perhaps no detectable activity at all), more often the variation in activity levels is continuously graded. Although the variation may be wide, and any particular individual may consistently show a level of activity characteristic of one region of the distribution rather than another, it is often not possible to classify individuals on a non-arbitrary basis into distinct groups because of the absence of any well defined discontinuities in the distribution of activities. Clearly defined bimodal or trimodal distributions of activities appear to be the exception rather than the rule, even when much of the variation is due to only a small number of genes. This makes precise genetical analysis difficult, unless some other more clear-cut criteria can be found to characterize the enzyme in individuals with different levels of activity.

The examples discussed below illustrate the manner in which some of these general problems have been, at least in part, resolved in the case of certain enzymes. They show in a variety of ways how single gene differences may lead to characteristic differences between individuals in enzyme activity. They also illustrate how the discrete and quite specific effects of a small number of different genes may underlie apparently continuously graded variation in activity levels.

5.2. Serum cholinesterase (acylcholine acyl hydrolase)

This enzyme is normally found in quite high activity in human blood serum. It is probably mainly synthesised in the liver. A variety of choline esters such as acetylcholine, butyrylcholine and benzoylcholine are readily hydrolysed as are a number of non-choline esters such as acetylsalicylic acid. Thus there is a fairly wide substrate specificity, and it is not known exactly what ester may serve as its natural substrate under normal conditions. The enzyme (sometimes also called pseudocholinesterase) must be distinguished from acetylcholinesterase (so-called 'true cholinesterase') which occurs in nervous

tissue and also in red cells and which has a much more restricted substrate specificity, particularly directed to acetylcholine.

The inherited variations of serum cholinesterase first came to light following the introduction and widespread use of the drug suxamethonium (succinyl dicholine) as a muscle relaxant in surgery and electroconvulsion therapy (fig. 5.1). Normally the effects of this drug are quite short because it is rapidly hydrolysed into inactive products by serum cholinesterase. How-

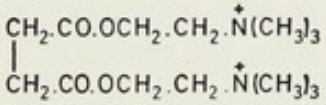


Fig. 5.1. Suxamethonium (succinyl dicholine).

ever occasional individuals (about 1 in 2000 in European populations) are unusually sensitive to its effects. Following a normal dose of the drug such people develop an extremely prolonged muscular paralysis and respiratory apnoea often lasting two hours or more, instead of the usual period of just a few minutes. It was found that in these individuals the level of serum cholinesterase was consistently low, and it seemed reasonable therefore to suppose that this was the cause of the sensitivity (Bourne et al. 1952, Evans et al. 1952). Furthermore a significant number of the immediate relatives of suxamethonium sensitive individuals were also found to have in various degrees reduced levels of serum cholinesterase activity (Lehmann and Ryan 1956). This indicated that the peculiarity might be genetically determined, and led to the suggestion that suxamethonium sensitive individuals with the low levels of serum cholinesterase are homozygous for an abnormal gene which in heterozygotes results in a moderate reduction in enzyme level. The three postulated genotypes could not however be clearly distinguished one from another on the basis of the levels of serum cholinesterase activity. In fact the distribution of serum cholinesterase levels in the population as a whole is effectively continuous, although the spread of values is very wide, and the suxamethonium sensitive individuals themselves show values at the lower end of the distribution.

5.2.1. 'Atypical' serum cholinesterase

An important advance was made when it was discovered that the serum cholinesterase present in suxamethonium sensitive individuals is atypical in certain of its properties (Kalow and Genest 1957, Kalow and Davies 1958, Davies et al. 1960). It was found for example, that when Michaelis constants are determined on the enzyme obtained from suxamethonium sensitive

people, using a number of different choline esters as substrates, the values obtained are for each substrate significantly greater than the corresponding values determined in unselected controls (Davies et al. 1960). Furthermore the magnitude of the difference varies from substrate to substrate (table 5.1).

TABLE 5.1

Michaelis constants for various choline esters with the 'usual' and 'atypical' serum cholinesterase (Davies et al. 1960).

	Michaelis constant (K_m) (mmole/l) 'Usual' enzyme	'Atypical' enzyme	Ratio of K_m 'atypical' : 'usual'
Acetylcholine	1.40 ±0.04	9.0 ±0.10	6.4:1
Propionylcholine	0.41 ±0.04	2.3 ±0.35	5.6:1
Butyrylcholine	0.29 ±0.03	1.2 ±0.08	4.1:1
Pentanoylcholine	0.72 ±0.04	1.5 ±0.17	2.1:1
Hexanoylcholine	0.57 ±0.09	0.82 ±0.06	1.4:1
Heptanoylcholine	0.38 ±0.22	1.11 ±0.14	2.9:1
Benzoylcholine	0.004±0.0003	0.022±0.003	5.5:1

Significant differences are also found with serum cholinesterase inhibitors of the type which contain a positively charged nitrogen atom either as a quaternary ammonium group (as in choline), or a substituted amino group (Kalow and Davies 1958). The enzyme present in suxamethonium individuals is less readily inhibited by each of these substances than the normal enzyme, and again there are differences in the magnitude of the effect from inhibitor to inhibitor (table 5.2). From these and other results (e.g. Clark et al. 1968) it became clear that in suxamethonium sensitive people the serum cholinesterase enzyme protein must be qualitatively different in structure from that which occurs in most other people. However it should be noted that in a number of properties, for example thermostability and electrophoretic mobility the two forms of the enzyme do not apparently differ.

The kinetic findings can be largely accounted for in terms of less efficient binding of the substrate or of the inhibitor by the atypical enzyme. Possibly

TABLE 5.2

Action of various inhibitors on 'usual' and 'atypical' serum cholinesterase (Kalow and Davies 1958).

Inhibitor	Log_{10} molar concentration giving 50% inhibition (pI_{50})		Ratio of concentrations giving 50% inhibition
	'Usual' enzyme	'Atypical' enzyme	I_{50} 'atypical' : I_{50} 'usual' (non-logarithmic)
Procaine	–4.36	–3.23	14:1
Decamethonium	–4.88	–2.74	141:1
Chlorpromazine	–5.36	–4.13	17:1
Dibucaine	–5.57	–4.27	20:1
Neostigmine	–6.89	–5.48	25:1
Physostigmine	–7.84	–6.57	18:1

the peculiarity lies in an altered configuration at the so-called 'anionic site' on the enzyme surface where the charged choline or choline-like grouping of the substrate or inhibitor is thought to become attached. However, whatever the structural nature of the peculiarity, the results from kinetic experiments indicate that the low levels of activity of the enzyme observed in suxamethonium sensitive individuals can be largely, if not entirely, accounted for in terms of its altered catalytic properties. The actual amount of the enzyme protein present in the serum of such people is probably not on average very different from that present in other individuals.

The inhibition studies led to the development of a rather simple test for detecting the atypical enzyme in routine studies. It involves the determination of the degree of inhibition of serum cholinesterase activity by the inhibitor dibucaine (fig. 5.2) under certain standard conditions (Kalow and Genest

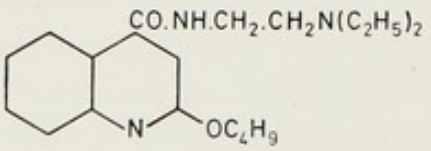


Fig. 5.2. Dibucaine (nupercaine, percaine).

1957). The percentage inhibition obtained is called the dibucaine number (DN), and has a value of about 20 ± 4 for the 'atypical' enzyme and about 80 ± 2 for the normal or 'usual' type enzyme (fig. 5.3). All individuals with

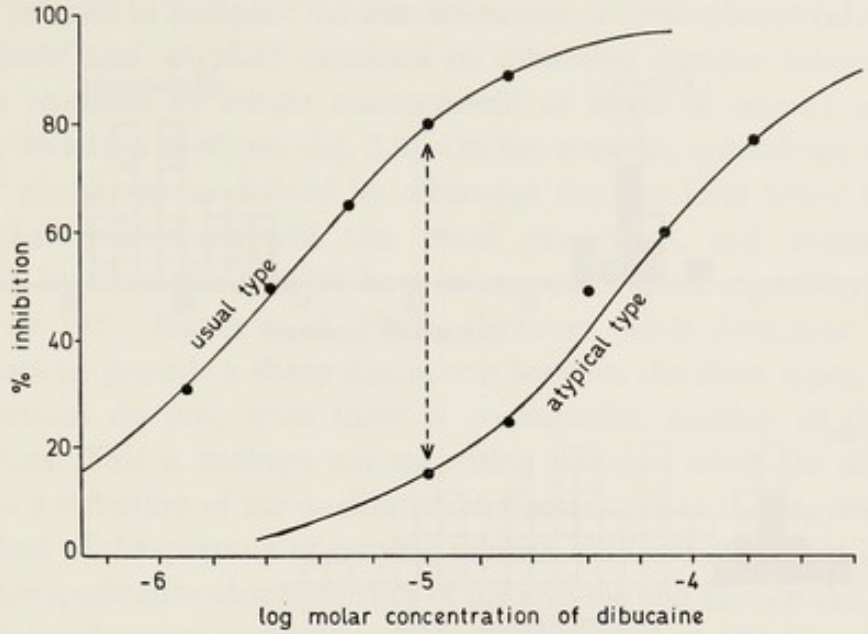


Fig. 5.3. Percentage inhibition of activity of the 'usual' and 'atypical' forms of serum cholinesterase by different concentrations of dibucaine (Kalow and Genest 1957). Optimal discrimination between the two forms of the enzyme is obtained with 10^{-5} M dibucaine. Activity is determined spectrophotometrically with benzoylcholine as substrate.

dibucaine numbers around 20 are extremely sensitive to suxamethonium.

When dibucaine number determinations are made on a random sample of the general population, a third group of people can also be clearly identified. They have dibucaine numbers of about 62 ± 4, and they are said to have the 'intermediate' phenotype. They occur with a frequency of about 1 in 25 in most European populations and they are not found in general to be particularly sensitive to suxamethonium. Such individuals appear to synthesise both the 'usual' and the 'atypical' forms of serum cholinesterase and do so in roughly equal amounts.

With the advent of dibucaine number determination detailed genetical studies became possible, and large numbers of families selected either through suxamethonium sensitive individuals with the 'atypical' enzyme or through individuals of the 'intermediate' type identified in population studies, were investigated (Kalow and Staron 1957, Harris et al. 1960, Harris and Whittaker 1962). Most of the results can be simply explained in terms of the segregation of two allelic genes, E_1^u which determines the 'usual' type enzyme, and E_1^a which determines the 'atypical' enzyme. Homozygotes $E_1^u E_1^u$ and $E_1^a E_1^a$ form only the corresponding enzyme. Heterozygotes $E_1^u E_1^a$ synthesise both types of enzyme and consequently give 'intermediate' dibucaine numbers.

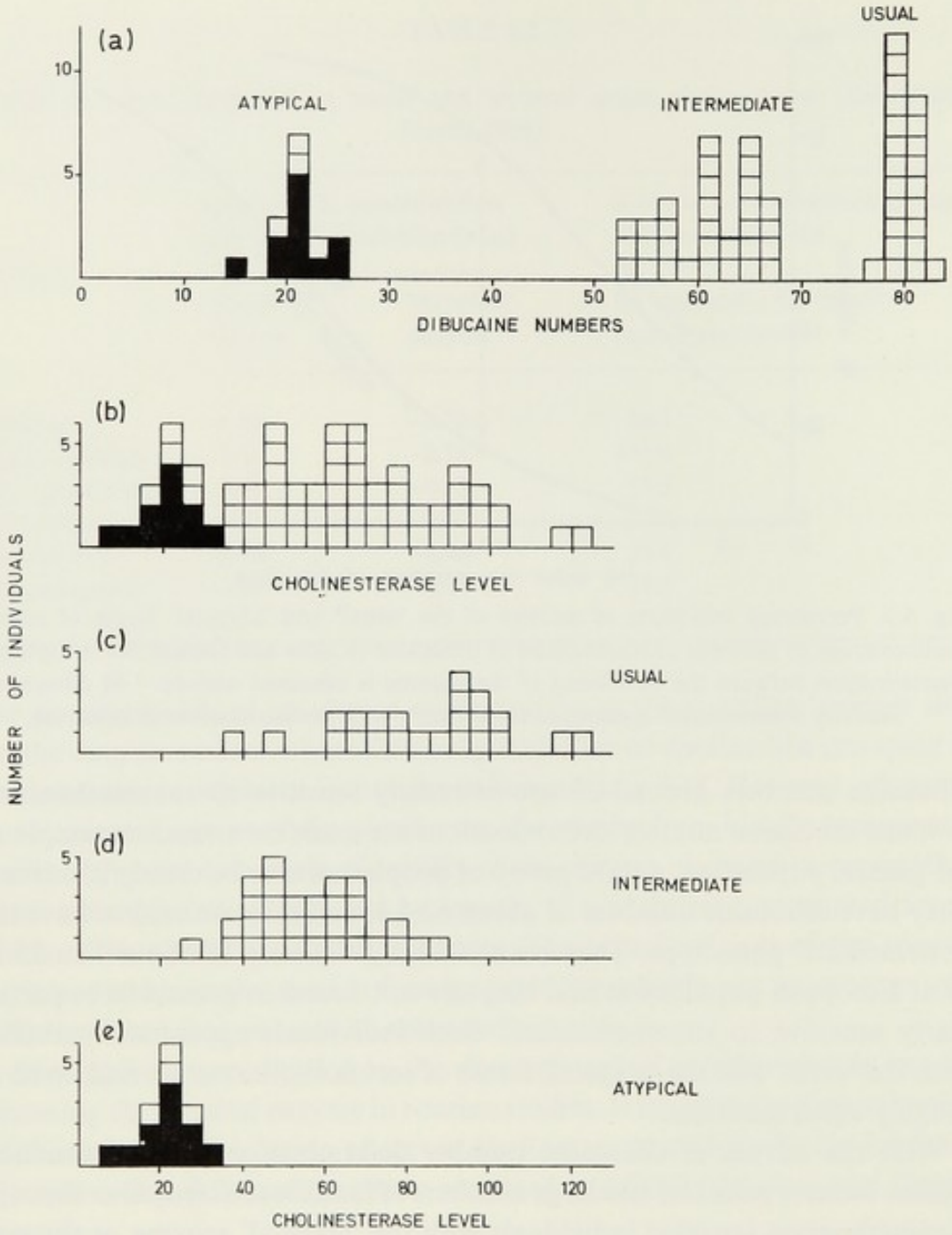


Fig. 5.4. Distributions of dibucaine numbers and of levels of cholinesterase activity in sera from 11 individuals found to be excessively sensitive to suxamethonium, and 58 of their relatives (Harris et al. 1960). Each square represents one individual, the suxamethonium sensitive propositi being marked in black. (a) Distribution of dibucaine numbers showing clear discrimination of the three phenotypes 'usual', 'intermediate' and 'atypical' (see text). (b) Distribution of levels of serum cholinesterase activity in the same serum samples as in (a). The activity levels were determined manometrically with acetylcholine as substrate. (c) Distribution of activity levels of those individuals classified as having the 'usual' phenotype in (a). (d) Distribution of activity levels of those individuals classified as having the 'intermediate' phenotype in (a). (e) Distribution of activity levels in those individuals classified as having the 'atypical' phenotype in (a). *Note:* Distribution (b) represents the sum of distributions (c), (d) and (e).

It is of interest to compare the discrimination of three phenotypes 'usual', 'intermediate' and 'atypical' obtained by dibucaine number determination with that obtained by simple measurement of levels of activity as determined by standard methods (fig. 5.4). On the average, individuals with the 'atypical' phenotype as defined by dibucaine number have lower levels of activity than individuals with the 'usual' phenotype, and 'intermediate' individuals by dibucaine number have intermediate levels of activity (Kalow and Staron 1957, Harris et al. 1960). However, while dibucaine number determinations provide a sharp distinction between the three types, activity determinations do not, since there is considerable overlap of the three distributions. This is perhaps not surprising because, while the dibucaine number is a reflection of the qualitative characteristics of the enzyme and is independent of the quantity of enzyme present, the level of activity depends both on the qualitative characteristics of the enzyme and also on the amount present. Since there are evidently many extraneous factors which can influence the amount of enzyme protein present in plasma at any given time, one would expect the activity determination to be a much more variable and less discriminating characteristic. In practice, one can find occasional individuals with the 'usual' type enzyme with levels of activity as low as those commonly found in the 'atypical' type. However, such individuals are very much less sensitive to suxamethonium and do not generally have a particularly marked apnoea.

5.2.2. The 'silent' allele

In the course of dibucaine number studies on families in which individuals of the 'atypical' type occurred, it became apparent that there are occasional exceptions to the rules of inheritance predicted by the simple two-allele hypothesis. If all individuals with only the 'atypical' enzyme are homozygous $E_1^a E_1^a$ then both their parents and all their children should also carry the gene E_1^a, and hence show either of the 'intermediate' or 'atypical' dibucaine number phenotypes. However, although this is usually the case, there are certain families in which one of the parents or children of an 'atypical' individual appear to have only the 'usual' type enzyme (Kalow and Staron 1957, Harris et al. 1960, Liddell et al. 1962, Simpson and Kalow 1964). Furthermore, non-biological parenthood as a general explanation for this phenomenon can be excluded.

The findings are in fact most simply accounted for by postulating the segregation in these exceptional families of a rare third allele, E_1^s, which is completely or almost completely ineffective in producing functionally active

enzyme. This has been referred to as the 'silent' allele. If such a gene occurred then one would expect that heterozygotes $E_1^aE_1^s$ would form only the 'atypical' enzyme, and heterozygotes $E_1^uE_1^s$ only the 'usual' enzyme (table 5.3).

TABLE 5.3

Serum cholinesterases formed by individuals with different combinations of alleles E_1^u, E_1^a and E_1^s

Genotype	Enzymes formed
$E_1^uE_1^u$ $E_1^uE_1^s$	'usual'
$E_1^uE_1^a$	'usual' + 'atypical'
$E_1^aE_1^a$ $E_1^aE_1^s$	'atypical'
$E_1^sE_1^s$	none

Also one would expect the occurrence of a further genotype $E_1^sE_1^s$ in which there is a complete or nearly complete absence of serum cholinesterase activity. A number of individuals with this rare condition have indeed been identified (Liddell et al. 1962, Doenicke et al. 1963, Hodgkin et al. 1965, Goedde et al. 1965). They are, as would be expected, extremely sensitive to suxamethonium.

One might also anticipate that $E_1^uE_1^s$ individuals would on the average have lower levels of activity than $E_1^uE_1^u$ individuals, and similarly $E_1^aE_1^s$ individuals would have lower levels of activity than $E_1^aE_1^a$ individuals. This has been found to be the case, but because there is considerable variation in serum cholinesterase level within each of these types it is not usually possible to identify them unequivocally through activity determinations alone.

In most populations the E_1^s allele is relatively uncommon, and so the genotype $E_1^sE_1^s$ is extremely rare. However in a somewhat isolated group of Eskimos living in Alaska, between 1% and 2% of the population have been found to show this virtually complete serum cholinesterase deficiency characteristic of the genotype $E_1^sE_1^s$ (Gutsche et al. 1967). Nearly 25% of individuals in this population group were evidently heterozygous $E_1^uE_1^s$. The E_1^a allele, however, was not found at all in this population. Apart from excessive sensi-

tivity to suxamethonium, $E_1^sE_1^s$ individuals who are virtually devoid of serum cholinesterase, are usually quite healthy.

5.2.3. 'Fluoride resistant' serum cholinesterase

A further variant of the enzyme was discovered in a rather unexpected way (Harris and Whittaker 1961). Serum cholinesterase is inhibited by low concentrations of sodium fluoride, and this is chemically so different from dibucaine and related substances which differentiate apparently very specifically between the 'usual' and 'atypical' forms of the enzyme, that it must be presumed to act in a quite different manner. It was therefore surprising to find that the two forms of the enzyme, 'usual' and 'atypical' can also be distinguished using fluoride as the inhibitor. The 'atypical' enzyme is much less readily inhibited than the 'usual' enzyme, and when a large number of different sera previously classified into 'usual', 'intermediate' and 'atypical' types by dibucaine number determination were examined using fluoride as inhibitor, the discrimination into the three phenotypes was, with a few exceptions, found to be virtually the same.

The occasional exceptions, however, turned out to be of particular interest. Family studies showed that they represented new serum cholinesterase phenotypes which could be correctly identified only by a combination of dibucaine and fluoride inhibitor tests (Harris and Whittaker 1962, Liddell et al. 1963, Whittaker 1967). It emerged that there is a further allele (E_1^f) which determines a third form of serum cholinesterase with properties different from the 'usual' and the 'atypical' forms previously recognised. This is referred to as the 'fluoride resistant' form. It may occur together with the 'usual' enzyme in individuals of genotype $E_1^uE_1^f$, or together with the 'atypical' enzyme in individuals of genotype $E_1^aE_1^f$. It may also occur alone as in the genotypes $E_1^fE_1^f$ and $E_1^fE_1^s$. The 'fluoride resistant' enzyme exhibits less activity than the 'usual' enzyme, but more than the 'atypical' enzyme, so the different types in which it occurs show on average differing levels of total serum cholinesterase activity.

5.2.4. Multiple allelism causing 'continuous' variation in activity

The four alleles (E_1^u, E_1^a, E_1^s and E_1^f) give rise to ten different genotypes which are listed in table 5.4. Their identification may require not only inhibitor tests with dibucaine and fluoride, but also detailed family studies. On average they differ from one another in the levels of activity they display, but there is considerable variation between individuals of any one type due

TABLE 5.4

Serum cholinesterase genotypes and phenotypes. Summary of inhibition characteristics with dibucaine (DN) and fluoride (FN) and of relative activities of serum cholinesterase in ten different genotypes. The values quoted are rounded off means based on the experience of several laboratories. The 'relative activities' are based on assays using benzoylcholine as substrate under standardised conditions. Such 'relative activities' will in general vary with the substrate and conditions used for the assay.

Genotype	Enzymes present	Inhibition characteristics		Relative activity
		DN	FN	
$E_1^uE_1^u$	'Usual'	80	60	100
$E_1^uE_1^f$	'Usual'+'fluoride resistant'	75	50	85
$E_1^uE_1^a$	'Usual'+'atypical'	60	50	75
$E_1^uE_1^s$	'Usual'	80	60	70
$E_1^fE_1^a$	'Fluoride resistant'+'atypical'	50	30	60
$E_1^fE_1^f$	'Fluoride resistant'	65	30	55
$E_1^aE_1^a$	'Atypical'	20	20	50
$E_1^fE_1^s$	'Fluoride resistant'	65	30	30
$E_1^aE_1^s$	'Atypical'	20	20	25
$E_1^sE_1^s$	—	—	—	0

to other factors. So the distribution of levels of activity in a large population containing the whole array of genotypes is effectively continuous.

Markedly prolonged paralysis following suxamethonium occurs as a regular phenomenon in three of these genotypes $E_1^aE_1^a$, $E_1^aE_1^s$ and $E_1^sE_1^s$. $E_1^aE_1^f$, $E_1^fE_1^s$ and $E_1^fE_1^f$ individuals also show some increase in sensitivity to the drug, though in different degrees.

The alleles E_1^u, E_1^a and E_1^f evidently determine structurally distinct forms of the enzyme protein. Possibly these only differ by single aminoacid substitutions, but so far it has not proved possible to characterise the differences precisely in structural terms. The enzymes differ in their kinetic properties with various substrates and inhibitors, but they appear to be very similar in other respects (e.g. electrophoretic mobilities and molecular size). The altered levels of activity observed in the different genotypic combinations of these alleles can probably be attributed largely to their different kinetics with the substrates used for the assay, rather than to any differences in rates of synthesis or in stability.

It is worth noting that the observed degree of reduction in the level of

activity in a particular genotype, e.g. $E_1^aE_1^a$, compared with that of the common or 'normal' genotype $E_1^uE_1^u$ may vary considerably according to the substrate used in the particular assay procedure. It will also be affected by the actual substrate concentration adopted for the assay and the other conditions. This is because the kinetics of these enzymes are complex, and kinetic variations from substrate to substrate do not necessarily parallel one another in the 'usual', 'atypical' and 'fluoride resistant' forms. This emphasises the complexity of such a parameter as 'level of enzyme activity' when it is used to compare different genotypes.

It is quite probable that several different mutants are included in the class of E_1^s alleles, and they may bring about the characteristic complete or almost complete loss of enzyme activity associated with this allele in different ways. In some cases it seems from immunochemical studies (Hodgkin et al. 1963, Goedde et al. 1965) that the absence or gross reduction of enzyme activity in homozygotes of genotype $E_1^sE_1^s$ is due to a true deficiency of enzyme protein. But it is not known exactly how this is brought about.

5.2.5. A second locus – E_2

There is yet a further source of genetically determined variation in serum cholinesterase activity. This was discovered by electrophoresis, which does not discriminate the various enzyme types discussed above (Harris et al. 1962).

When serum cholinesterase is examined electrophoretically several isozymes (known as C_1, C_2 and C_3 and C_4) are regularly observed. C_1, C_2 and C_3 are minor components and contribute little to the total activity, most of which is derived from C_4 (fig. 5.5). However in about 10% of people in European and certain other populations, an extra isozyme called C_5 is seen as well (Harris et al. 1962, 1963a; Simpson 1966, Robson and Harris 1966). The presence of C_5 in serum appears to be determined by a gene at a locus which is separate and not closely linked to the locus at which the alleles E_1^u, E_1^a etc. occur (Harris et al. 1963c, Simpson 1966). The amount of serum cholinesterase activity attributable to C_5 varies considerably from individual to individual carrying the appropriate gene. However sera from C_5^+ individuals (i.e. those with the extra isozyme) have on average about 25% higher levels of activity than sera from those without it (so-called C_5^- individuals). But the variation is wide in both groups and the two distributions overlap to a very considerable extent.

The mode of action of the gene determining the C_5 component is not understood. It apparently leads to the appearance of an electrophoretically distinct isozyme whose properties are in other respects similar to those of the

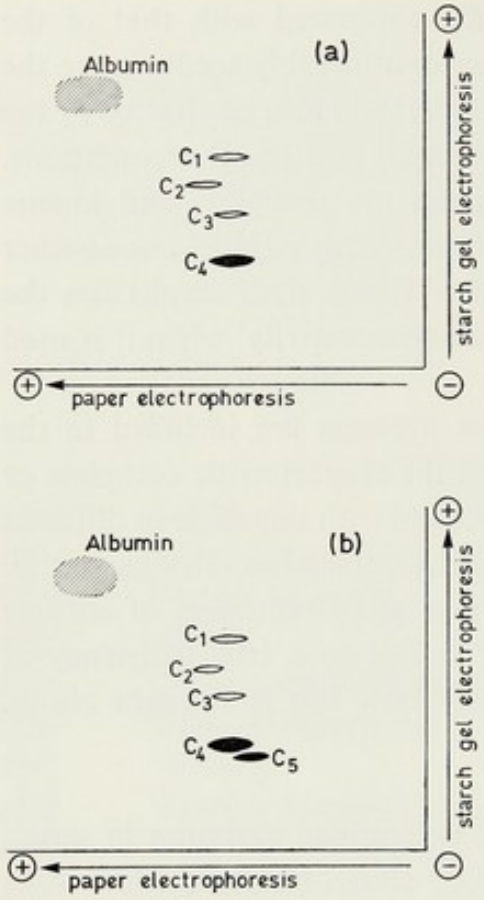


Fig. 5.5. Diagrams showing separation of serum cholinesterase isozymes by two dimensional filter paper/starch gel electrophoresis at *p*H 8.6 (Harris et al. 1962). (a) C_5- serum; (b) C_5+ serum. Isozyme C_4 is present in both (a) and (b), and accounts for most of the serum cholinesterase activity in (a) and a high proportion (ca. 65-90%) in (b). Isozymes C_1, C_2, and C_3 are minor components present in both (a) and (b). Isozyme C_5 is the additional component present in (b). Note that the electrophoretic mobilities towards the anode in starch gel are in the order $C_1>C_2>C_3>C_4>C_5$ but in filter paper the order is $C_2>C_1 = C_3 = C_4>C_5$.

This is because the isozymes differ from each other both in molecular charge and also in molecular size. The molecular sizes of the various isozymes are in the order $C_4 \sim C_5 > C_3 > C_2 > C_1$ (Harris and Robson 1963).

other serum cholinesterase isozymes, and which appears to be affected in the same way by the alleles E_1^u, E_1^a etc.

Thus the genetical factors concerned in determining the level of serum cholinesterase are complex. At least two chromosomal loci are involved. At one locus (E_1) alleles determining both qualitative and quantitative differences in the enzyme occur. At the other locus (E_2) one allele E_2^+ determines the formation of an extra serum cholinesterase isozyme C_5, while the other allele (E_2^-) appears to be functionally inactive. The level of activity observed in any one person will depend on the particular alleles present at these two loci as well as no doubt on a variety of extraneous non-genetical factors, both in healthy individuals and in patients with particular types of disease. The level of serum cholinesterase for example is usually depressed in liver disease, presumably because of diminished formation of the enzyme protein.

5.3. Glucose-6-phosphate dehydrogenase (G-6-PD)

5.3.1. G-6-PD deficiencies

The discovery of inherited variation of glucose-6-phosphate dehydrogenase (G-6-PD) followed the finding that a significant proportion of American Negroes develop an acute haemolysis when they receive the synthetic anti-malarial drug, primaquine (Hockwald et al. 1952). It was shown that the haemolytic response to the drug was due to an intrinsic red cell abnormality (Dern et al. 1954), and that this was a specific deficiency of the enzyme glucose-6-phosphate dehydrogenase (Carson et al. 1956).

Glucose-6-phosphate dehydrogenase catalyses the oxidation of glucose-6-phosphate to 6-phosphogluconate, and this is accompanied by the concomitant reduction of the coenzyme NADP to $NADPH_2$. The reaction represents the first step in the oxidation of glucose via the so-called pentose shunt pathway (see fig. A1 p. 267). Under normal conditions this pathway accounts for only a small proportion of the glucose utilised by the red cell. It does however serve to maintain the intracellular concentration of the reduced coenzyme $NADPH_2$, and it seems likely that it is a failure in this which accounts through other reactions, particularly the maintenance of glutathione in the reduced form, for the haemolytic crises observed when G-6-PD deficient subjects receive primaquine and other drugs.

Assays of red cell G-6-PD levels in Negro populations reveal a striking sex difference (fig. 5.6). In healthy males two clearly distinct classes of individuals can be recognised; those with normal levels of the enzyme, and those who are deficient. Although in each group there is considerable variation from one individual to another, the two distributions hardly overlap. The deficient group has levels of red cell G-6-PD varying around 15% of the average found in the non-deficient group. A quite different situation is found in females. Here a more or less continuous variation in levels of the enzyme in different individuals is observed. All values may be found from the normal levels characteristic of the non-deficient males to the low levels characteristic of the deficient males.

The significance of the sex differences became clear when family studies showed that the variations are determined by alleles at a locus on the X chromosome (Childs et al. 1958). Males have only one X chromosome and so may either carry the normal allele or a mutant allele causing the deficiency, but not both. Females have two X chromosomes. They may be homozygous for the normal allele and have normal levels of the enzyme like normal males; homozygous for the mutant allele and have a marked deficiency of the

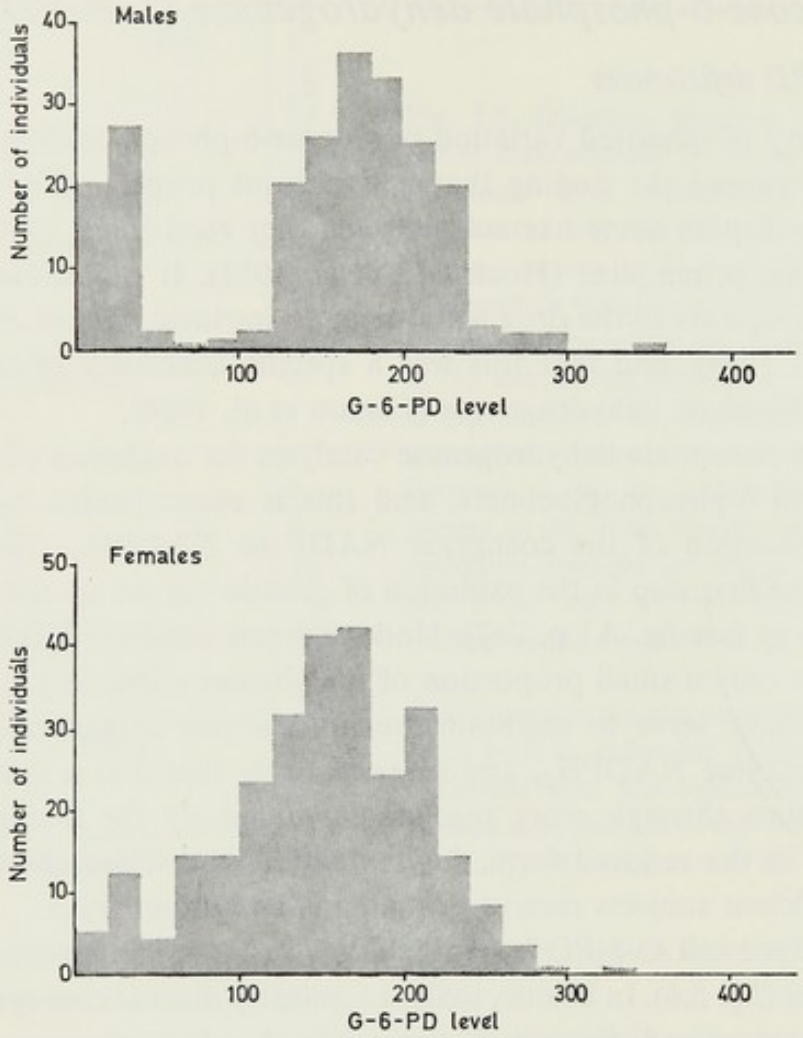


Fig. 5.6. Distribution of red cell glucose-6-phosphate dehydrogenase activity levels in 235 male and 284 female Nigerians (data of Nance 1967).

enzyme like deficient males; or heterozygous, in which case they have levels of the enzyme which are on the average almost exactly intermediate between the average level of the normal males and the average level of the deficient males. Since considerable variation in level of enzyme activity among heterozygous females occurs, the distribution of values overlaps at one extreme with the distribution of homozygous deficient values, and at the other with the distribution of homozygous normal values. Consequently it is often not possible to classify the female unequivocally into one or other of the three postulated genotypes simply in terms of the level of the enzyme in red blood cells.

It was soon recognised that other drugs besides primaquine may induce haemolytic crises in G-6-PD deficient subjects (table 5.5). These include sulfonamides such as sulfapyridine and sulfanilamide, other antibacterial

TABLE 5.5

Drugs and other agents that can cause clinically significant haemolysis in G-6-PD deficient people (from WHO Technical Report Series No. 366, 1967).

Acetanilid	Nitrofurazone (Furacin)
Phenylhydrazine	Nitrofurantoin (Furadantin)
Sulfanilamide	Furazolidone
Sulfacetamide	Furaltodone (Altofur)
Sulfapyridine	Quinidine
Sulfamethoxypyridazine (Kynex)	Primaquine
Salicylazosulfapyridine (Azulfidine)	Pamaquine
Thiazosulfone	Pentaquine
Diaminodiphenylsulfone	Quinocide
Trinitroluene	Naphthalene
	Neosalvarsan
	Fava Beans

agents such as nitrofurantoin, other antimalarials such as pentaquine, and so on. It was also discovered that a form of G-6-PD deficiency is the underlying cause of the disease known as favism, which has long been recognised as a not infrequent condition in certain population groups living in the Middle East and some Mediterranean countries (Zinkham et al. 1958, Larizza et al. 1958, Szeinberg et al. 1958). It manifests as an acute haemolytic anaemia which follows the ingestion of fava beans, a not uncommon feature of the diet in these areas.

5.3.2. The common Negro and Mediterranean G-6-PD variants: Gd B, Gd A, Gd A-, and Gd Mediterranean

Like the form of G-6-PD deficiency occurring in Negroes, the G-6-PD deficiency found in Mediterranean and Middle East populations is determined by a mutant gene on the X chromosome. A different mutation is however involved, and there are certain characteristic differences between the two conditions. In particular the level of the enzyme in red cells of affected males with the so-called 'Mediterranean' type of G-6-PD deficiency is usually about 3 or 4% of the normal level, whereas in the 'Negro' type it is on average about 15%. Also a significant reduction in G-6-PD level may be demonstrated in white cells (Ramot et al. 1959) and in other tissues in the Mediterranean type, whereas in the Negro type such a reduction is either not

found or is very slight (Marks et al. 1959). Thus in terms of the level of enzyme activity the Mediterranean type is a relatively more severe abnormality.

Another difference between the two kinds of deficiency was brought to light by electrophoretic studies. These studies also revealed a further kind of variation in G-6-PD which is relatively common in Negro populations, but is not associated with marked enzyme deficiency (Boyer et al. 1962, Kirkman and Hendrickson 1963).

When G-6-PD from different healthy Negro subjects is examined by electrophoresis, two separate forms may be separated (fig. 5.7). The slower

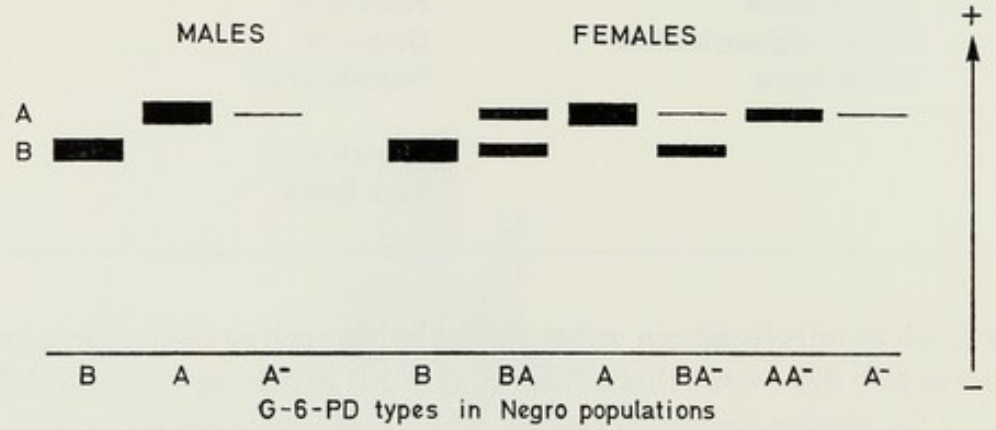


Fig. 5.7. Diagram showing electrophoretic components of red cell G-6-PD in Negroes of different phenotypes.

moving form at *p*H 8.6 is called B and the faster one A. Males may have an enzyme with the B mobility or with the A mobility but not both. Furthermore in those who are G-6-PD deficient the enzyme present almost invariably shows the A mobility. Thus three distinct classes of Negro male, each of which is relatively common, may be distinguished from the level of G-6-PD activity present in the red cells and from its electrophoretic properties. These three phenotypes are referred to as Gd B, Gd A and Gd A-. Gd B individuals show normal levels of G-6-PD activity and the B electrophoretic mobility. Gd A individuals show normal levels of activity and the A electrophoretic mobility. Gd A- individuals show deficient enzyme activity (about 15% of normal) and A electrophoretic mobility.

The relative frequencies of these three phenotypes vary somewhat among different Negro populations. For example, among male Nigerians (Yoruba) about 56% were found to be Gd B, about 22% were Gd A, and about 22% were Gd A- (Porter et al. 1964). Among male Negroes living in the U.S.A.

typical figures are Gd B: 60–70%, Gd A: 15–20%, Gd A-: 10–15%. Family studies indicate that these phenotypes are determined by three distinct alleles occurring at a single locus on the X chromosome. They have been designated Gd^B, Gd^A and Gd^{A-}. Among Negro females six different genotypes occur, and the corresponding phenotypes are illustrated diagrammatically in fig. 5.7. Three of these are similar to the three phenotypes seen in males and correspond to the homozygous genotypes Gd^BGd^B, Gd^AGd^A and $Gd^{A-}Gd^{A-}$. The other three correspond to the heterozygous genotypes Gd^BGd^A, Gd^BGd^{A-} and Gd^AGd^{A-}.

In the so-called Mediterranean type of G-6-PD deficiency, the enzyme which is present has the B electrophoretic mobility, as does the enzyme in normal individuals in these populations. The allele causing the Mediterranean type of deficiency is almost certainly at the same locus as the common alleles occurring in Negro populations. It has been designated $Gd^{Mediterranean}$.

The four alleles, Gd^B, Gd^A, Gd^{A-} and $Gd^{Mediterranean}$, appear to determine four structurally distinct forms of the G-6-PD enzyme protein (see table 5.6, pp. 128–129). The so-called B enzyme must be regarded as the normal type, because it is by far the commonest and occurs in all populations. The others possibly differ from it simply by single aminoacid substitutions. Evidence on this point has been obtained from structural studies in the case of the A enzyme (determined by the Gd^A allele). Here it appears that a particular asparagine residue present in the B enzyme is replaced by an aspartic acid residue (Yoshida 1967). So far it has not proved possible to carry out similar structural investigations on the Gd A- or the Gd Mediterranean types of the enzyme because of the difficulty of isolating them in pure form in adequate amounts. However, comparisons of the properties of these various enzymes have revealed a number of other interesting differences and similarities.

The A enzyme, although it differs from the normal B enzyme electrophoretically, has been found to resemble it closely in many other respects. The two enzymes appear to have essentially the same K_m for G-6-P and NADP, similar *p*H optima, and similar rates of thermal denaturation. Thus the structural alteration, while causing an electrophoretic difference, does not apparently lead to any marked differences in catalytic activity or in stability. There may however be minor differences not detectable by present techniques because assays of G-6-PD activity in red cells in large numbers of Gd B and Gd A individuals have shown that there is small difference in the average level of activity. The mean level in Gd A individuals appears to be about 10% less than in Gd B individuals, though there is a very considerable overlap in the two distributions.

The A- enzyme (determined by the Gd^{A-} allele) has the same electrophoretic mobility as the A enzyme and is very similar both to it and the B enzyme in a variety of kinetic and other properties (Kirkman 1959, Marks et al. 1961). It differs markedly, however, in its stability *in vivo*. This has been demonstrated by comparing enzyme activities in relatively 'young' and relatively 'old' red cells (Yoshida et al. 1967, Piomelli et al. 1968). The cells were fractionated into 'age groups' according to their specific gravity by centrifugation in a density gradient. It was found that in the younger cells (with a high proportion of reticulocytes) the level of G-6-PD activity in Gd A- individuals is almost as high as in Gd B individuals. However, in relatively old cells there is a very marked difference, the older cells from Gd B individuals showing much more G-6-PD activity than comparable cells from Gd A- individuals. It has been estimated from these studies that the half life of the enzyme in red cells from normal Gd B individuals is about 62 days, whereas in Gd A- individuals the half life of the enzyme in red cells is only about 13 days (Piomelli et al. 1968). So the enzyme deficiency observed in Negro Gd A- individuals is evidently due to the more rapid denaturation of the A- enzyme protein *in vivo*. A sample of red cells used in the ordinary assay of the enzyme in an individual consists of course of a mixture of cells of all ages. The level of activity seen in normal Gd B individuals or in Gd A- individuals represents an average of the relatively higher values present in the younger cells, and the lower values in the older ones.

This age differential largely explains why no marked deficit of G-6-PD activity is seen in white cells from Gd A- individuals. Red cells normally have a life span of more than 100 days, but they lose their nuclei and their ability to synthesise proteins at an early stage. Consequently the G-6-PD enzyme protein, which is progressively reduced in quantity by denaturation, is not replaced by newly synthesised enzyme. But white cells which are nucleated are capable of continuing enzyme synthesis. Thus the effect of the more rapid decay of the A- enzyme is much less pronounced in circulating white cells than in circulating red cells, and may indeed be barely detectable.

The Gd Mediterranean enzyme, although it has the same electrophoretic mobility as the B enzyme, differs in a number of other properties (Kirkman et al. 1964a). The K_m values for G-6-P and also for NADP are both appreciably lower than the corresponding values obtained for the B enzyme. Another difference is in the rate of denaturation at elevated temperatures, which is significantly increased in Gd Mediterranean. There are also differences in pH optima. A particularly striking peculiarity is in its ability to utilise 2-deoxy-glucose-6-phosphate at about 25–30% of the rate at which it oxidises G-6-P.

The distinctive peculiarities in the properties of the enzyme presumably reflect in different ways its altered structure. They do not, however, *per se* account for the very low levels of activity seen in the Gd Mediterranean phenotype. This is evidently due, as in the case of the A-enzyme, to a decreased stability of the enzyme protein *in vivo*. But here the effect is even more profound. The rate of decay is evidently extremely rapid, and even though younger red cells show a somewhat less severe deficiency of the enzyme than older cells, the reduced level of enzyme activity, even in reticulocytes, is nevertheless quite marked (Piomelli et al. 1968). Also a significant G-6-PD deficit is found in white cells in Gd Mediterranean individuals (Ramot et al. 1959).

The Gd A- type of G-6-PD deficiency is, as we have seen, relatively common among Negroes, and the Gd Mediterranean type is relatively common in non-Negro populations living in the Middle East and Southern Europe. Most individuals affected with either of these forms of enzyme deficiency are quite healthy and do not show any ill effects unless they happen to take one or other of the various drugs to which they are sensitive or, in the case of the Gd Mediterranean type, unless they happen to eat fava beans. Under these circumstances they are liable to develop an acute haemolytic reaction. But the severity of this reaction is very variable. Thus some Gd Mediterranean individuals appear to be able to eat the fava bean without serious consequences while others, after similar amounts, show a rapid and severe response. The cause of these variations is not known. They may in part be genetically determined (Stamatoyannopoulos et al. 1966) and could perhaps have to do with differences in the manner in which the active principle of the bean is metabolised.

5.3.3. Other G-6-PD variants

Besides these two standard types of G-6-PD deficiency a number of other forms have been identified in a variety of different population groups. They each appear to be due to the synthesis of a structurally distinctive form of the enzyme with characteristic properties. Most of the alleles which determine them are relatively rare and appear to be irregularly distributed in different populations. But certain alleles occur with an appreciable frequency in particular population groups. For example, an allele which gives rise to the variant known as Gd Canton is relatively common in populations originating from around Southern China, and Gd Athens is relatively frequent in Greece.

Each of these G-6-PD variants differs from one another and from the Gd B, Gd A, Gd A- and Gd Mediterranean forms of the enzyme, in one or another of their qualitative characteristics (e.g. electrophoretic

TABLE 5.6

Variants of glucose-6-phosphate dehydrogenase.

Variant	Red cell activity (% of normal)	Electrophoretic mobility relative to normal (*p*H 7.0 or *p*H 8.6)	K_m for G-6-P (μM)	K_m for NADP (μM)
Normal (Gd B)	100	—	50-78	2.9-4.4
1. Hektoen	400	fast (*p*H 6.5)	n	n
2. Madison	100	slow	?	?
3. A	90	fast	n	n
4. Baltimore-Austin	75	slow	n	n
5. Madrona	70-80	slow	reduced (32)	n
6. Ibadan-Austin	72	slow	n	n
7. Barbieri	40-60	fast	increased	increased
8. Kerala	50	slow	reduced (23)	reduced (1.5)
9. Tel-Hashomer	25-40	slow	reduced (30-40)	?
10. Athens	25	slow	reduced (16-19)	n
11. *Chicago	9-26	normal	n	n
12. Seattle	8-21	slow	reduced (15-25)	reduced (2.4-2.8)
13. A-	8-20	fast	n	n
14. Canton	4-24	fast	reduced (20-36)	reduced (2.0-2.4)
15. West Bengal	9	slow	reduced (31)	increased (6.6)
16. * Ohio	2-16	fast	slightly increased	slightly increased
17. * Oklahoma	4-10	normal	increased (127-200)	increased (20)
18. * Duarte	8.5	normal	n	n
19. Mediterranean	0-7	normal	reduced (18-26)	reduced (1.2-1.6)
20. * Alberquerque	1	normal	increased (115)	increased (11)
21. * Eyssen	0	slow	?	?

* variant associated with congenital haemolytic disease.

Utilisation of substrate analogue 2-dG-6-P relative to G-6-P	Thermostability relative to normal	*p*H activity curve	Incidence	References
<4%	—	truncate	Commonest type in all populations	
n	n	n	rare	Dern et al. (1969)
?	?	?	rare	Nance and Uchida (1964)
n	n	n	Common in Negroes	Boyer et al. (1962) Kirkman et al. (1964a)
n	n	n	rare	Long et al. (1965)
n	?	n	rare	Hook et al. (1968)
n	n	n	rare	Long et al. (1965)
?	n	?	rare	Marks et al. (1962)
increased (7.4)	n	biphasic	rare	Azevedo et al. (1968)
n	n	slightly biphasic	rare	Ramot and Brok (1964)
increased (10-15)	slightly reduced	slightly biphasic	Common in Greece	Stamatoyannopoulos et al. (1967)
n	much reduced	n	rare	Kirkman et al. (1964b)
increased (7-11)	n	slightly biphasic	rare	Kirkman et al. (1965)
n	n	n	Common in Negroes	Boyer et al. (1962), Kirkman et al. (1964a)
increased (4-15)	slightly reduced	biphasic	Common in S.E. Asia	McCurdy et al. (1966)
n	n	n	rare	Azevedo et al. (1968)
n	much reduced	?	rare	Pinto et al. (1966)
n	reduced	narrow peak	rare	Kirkman et al. (1964a)
increased (5.4)	much reduced	narrow peak	rare	Beutler et al. (1968)
increased (23-27)	reduced	biphasic	Common in Mediterranean countries and Middle East	Kirkman et al. (1964c)
n	much reduced	narrow peak	rare	Beutler et al. (1968)
?	much reduced	?	rare	Boyer et al. (1962)

n: characteristic within normal range

mobility, K_m, thermostability, utilisation of 2-deoxy G-6-P, *p*H activity curve etc.). Some of these characteristics are summarised in table 5.6, and it is apparent that quite a variety of different properties must be examined for the positive identification and characterisation of a particular variant.

Many of these variants are associated with some degree of enzyme deficiency, but the cause of this probably differs in different cases. In some, decreased stability of the enzyme protein is probably the main cause, as it is with the Gd A- and the Gd Mediterranean variants. In others the low activity may be mainly due to alteration of the catalytic efficiency of the enzyme caused by its altered structure. In still other cases the deficit may be due to a reduced rate of synthesis of the enzyme protein, analogous perhaps to the decreased rate of synthesis of haemoglobin found in the thalassaemias. It is also possible, of course, that more than one of these different sorts of effect may be important in determining the level of activity observed in any specific instance.

It will be seen from table 5.6 that although many of the variants are associated with a very marked deficiency of the enzyme activity, with others the reduction in activity is only moderate, and some apparently lead to no significant alteration in level of activity at all. Furthermore, at least one variant (Gd Hektoen) is associated with a marked elevation in activity. Thus the different variants that have been characterised thus far encompass a very wide range of activity levels.

It will also be noted that certain variants are associated with a chronic form of haemolytic anaemia (so-called non-spherocytic haemolytic disease) which is present even in the absence of any obvious precipitating factors such as a particular drug, a dietary ingredient like the fava bean, or an infection. Chronic haemolytic disease, however, is not found in association with a number of other variants (e.g. Gd Mediterranean) even though the level of enzyme activity in the red cell may be apparently reduced to as low a level or even lower. Such apparent anomalies are probably in most cases attributable to the particular kinetic characteristics of the enzymes, and the manner in which these affect red cell function *in vivo*. For example, with conventional assay procedures haemolysates of Gd Oklahoma subjects show slightly greater activity than those of Gd Mediterranean subjects (table 5.6). Yet the former abnormality is associated with chronic haemolytic disease and the latter is not. The probable explanation for the difference is that the K_m of Gd Oklahoma with glucose-6-phosphate is very much greater than the corresponding K_m of Gd Mediterranean. In the standard assay system the concentration of glucose-6-phosphate is kept relatively high so as to saturate

the enzyme and ensure maximal reaction velocities. However, in the red cell the concentration of glucose-6-phosphate is very much less and is probably close to or lower than the K_m values. As can be seen from fig. 5.8 (Kirkman

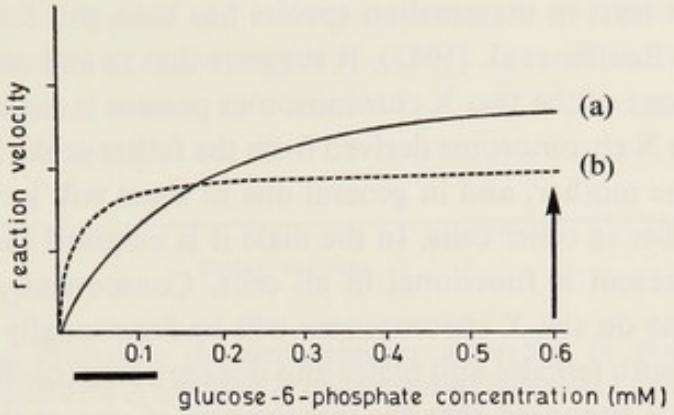


Fig. 5.8. Plots of reaction velocities at different glucose-6-phosphate concentrations of two G-6-PD variants (Kirkman 1968). (a) Gd Oklahoma (K_m = 140 μM) – continuous line; (b) Gd Mediterranean (K_m = 18 μM) – broken line. The curves are constructed so that the relative rates with standard assay concentrations of glucose-6-phosphate (indicated by the arrow) are proportional to the specific activities of haemolysates of the two variants. The solid bar indicates the actual range of intracellular concentration of glucose-6-phosphate.

1968), the reaction rate achieved in the red cell by Gd Oklahoma may well be very much less than that achieved by Gd Mediterranean, even though the relative levels of activity as determined in an assay system with high glucose-6-phosphate concentrations are in the opposite direction.

5.3.4. G-6-PD and the Lyon hypothesis

Since the locus determining G-6-PD is on the X chromosome it is represented twice in cells of the female (XX) but only once in cells of the male (XY). Nevertheless on the average the level of G-6-PD activity found in normal females is much the same as that in normal males (Marks 1958). Furthermore the same average level of activity is observed in individuals who for one reason or another possess abnormal numbers of X chromosomes e.g. males with the Klinefelter syndrome XXY and females with the XXX syndrome, provided that they do not also carry one of the alleles determining G-6-PD deficiency (Grumbach et al. 1962, Harris et al. 1963b). Thus in general the number of G-6-PD genes present in an individual does not appear to influence the level of activity obtained.

This phenomenon which appears to involve the equalisation of the effects

of X linked genes in the two sexes is sometimes referred to as dosage compensation. It probably occurs with respect to most loci on the X chromosome which are not concerned with characteristics peculiarly relevant to sexual differentiation. A hypothesis aimed at providing a general explanation of this phenomenon at least in mammalian species has been put forward by Lyon (1962) and also Beutler et al. (1962). It suggests that in any cell of the female organism only one of the two X chromosomes present is functionally active. This may be the X chromosome derived from the father or the X chromosome derived from the mother, and in general one of these will be active in some cells and the other in other cells. In the male it is assumed that the single X chromosome present is functional in all cells. Consequently only a single dose of any gene on the X chromosome will be functionally active in each somatic cell of both females and males and dosage compensation would thus be achieved. It is supposed that the inactivation of one or the other of the two X chromosomes present in the cell of the female occurs more or less at random and at a fairly early stage of embryological development. It is also supposed that once it has occurred in any one cell the same X chromosome continues to be inactivated in all the daughter cells subsequently derived from it. Thus according to this hypothesis a female may be regarded as a mosaic. In approximately half of her cells only the X chromosome derived from her father will be functionally active, while in the other cells only the X chromosome of her mother is active.

A direct way of testing the hypothesis is to consider the situation in a female heterozygous for two alleles at a locus on the X chromosome which determines the structure of a particular enzyme or protein. In some cells the functionally active chromosome should carry one allele and in others the other allele, so that some cells should synthesise one form of the enzyme or protein while the other cells should only synthesise the other form. A single cell should not contain both sorts of enzyme or protein. Investigation of heterozygotes for alleles determining G-6-PD variants have shown that this in fact appears to be the case.

A particularly convincing demonstration of the phenomenon was obtained in experiments using cells grown in tissue cultures started from small explants of skin from heterozygous Negro females of G-6-PD genotype $Gd^{B}Gd^{A}$ (Davidson et al. 1963). When the G-6-PD present in clones derived from individual cells in the culture were examined electrophoretically it was found that some clones showed only the A enzyme, and others only the B enzyme but none showed both (fig. 5.9). Furthermore in a tissue culture derived from skin from a single heterozygote both A clones and B clones could be obtained.

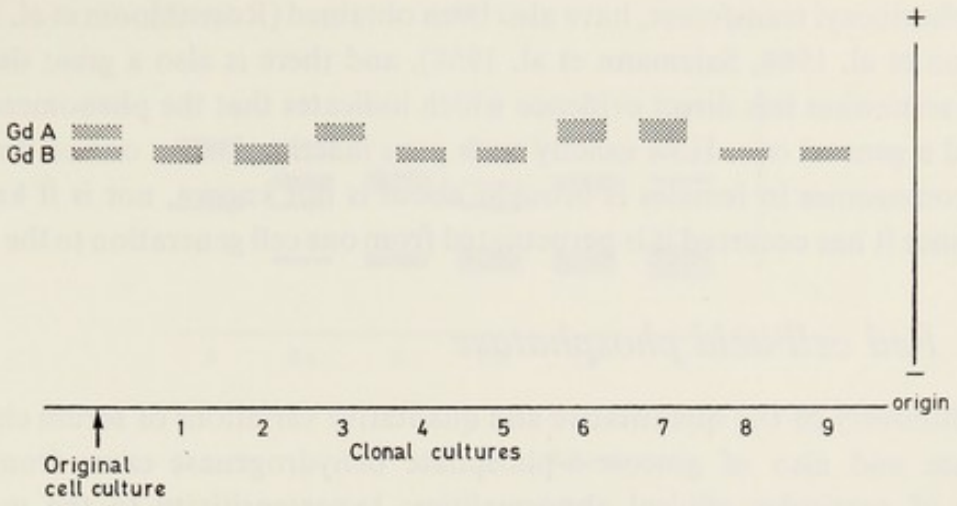


Fig. 5.9. Diagram illustrating the electrophoretic separation of glucose-6-phosphate dehydrogenase components in sonicates of tissue cultured cells from a heterozygous female of genotype $Gd^{A}\,Gd^{B}$ (redrawn from Davidson et al. 1963). The original cell culture shows the two G-6-PD components Gd A and Gd B. However the clones derived from single cells of this culture show either Gd A or Gd B but not both.

Thus it appeared that two distinct cell populations were present, one synthesising only the A enzyme and the other only the B enzyme.

Similar experiments were also carried out in tissue cultures derived from females heterozygous for the Mediterranean type of G-6-PD deficiency genotype $Gd^{B}Gd^{Mediterranean}$. Here it was found that two types of clone could be clearly distinguished. In one the level of enzyme activity was essentially the same as in tissue culture cells derived from normal males of genotype Gd^{B}, and in the other the level of enzyme activity was grossly reduced and equivalent to that found in tissue culture cells obtained from males of genotype $Gd^{Mediterranean}$. Thus again two biochemically distinct cell populations were present and both could be identified in tissue cultures derived from a single heterozygous individual.

Studies carried out on red cells from Negro females heterozygous for G-6-PD deficiency ($Gd^{B}Gd^{A-}$) indicate that here also two biochemically distinct populations probably occur (Beutler and Baluda 1964). Evidence pointing to the same general conclusion has also been obtained by studies on the electrophoretic patterns of the enzyme derived from small skin biopsies and from single tumours (leiomyomas) in heterozygotes of G-6-PD genotype $Gd^{B}Gd^{A}$ (Linder and Gartler 1965a, b).

Thus in the case of the G-6-PD locus the postulate that in any one cell of the female only one of the two alleles present is functionally active appears to be correct. Similar results for another X linked enzyme, hypoxanthine-guanine

phosphoribosyl transferase, have also been obtained (Rosenbloom et al. 1967, Migeon et al. 1968, Salzmann et al. 1968), and there is also a great deal of other somewhat less direct evidence which indicates that the phenomenon is indeed a general one. How exactly such gene inactivation on one of the two X chromosomes in females is brought about is not known, nor is it known how once it has occurred it is perpetuated from one cell generation to the next.

5.4. Red cell acid phosphatase

The discovery of the quantitative and qualitative variations of serum cholinesterase and also of glucose-6-phosphate dehydrogenase came from the study of particular clinical abnormalities; hypersensitivity to the muscle relaxant drug suxamethonium in the one case, and drug induced haemolytic anaemia and favism in the other. These studies led to the recognition of specific enzyme deficiencies which were subsequently shown to be due to the occurrence of qualitatively distinct variant forms of the particular enzyme. But in the case of red cell acid phosphatase, the recognition of inherited quantitative variation stemmed from a direct search for commonly occurring and structurally distinct variants of the enzyme among randomly selected healthy individuals (see pp. 225–227).

The enzyme is a phosphohydrolase and phosphotransferase with a low *p*H optimum which appears to be peculiar to the red cell, since it differs in a variety of ways from the acid phosphatases found in other tissues. However, its precise metabolic function in the red cell is not known. Nor has any particular clinical abnormality been found to be associated with a specific defect of the enzyme.

5.4.1. The electrophoretic variants

When acid phosphatase in red cells from different individuals is examined by electrophoresis, two or more distinct electrophoretic components or isozymes are regularly observed, and there are clear cut person to person differences in isozyme pattern (Hopkinson et al. 1963, 1964). Six different phenotypes can be recognised in European populations. They are referred to as types A, BA, B, CA, CB and C and the characteristic electrophoretic patterns of red cell acid phosphatase in these different types are illustrated in fig. 5.10. Among Europeans about 13% of people are type A, 43% type BA, 36% type B, 3% type CA, 5% type CB and about 1 in 600 people are type C.

Family studies show that the different types are genetically determined, and the segregation of the types among offspring of the various possible

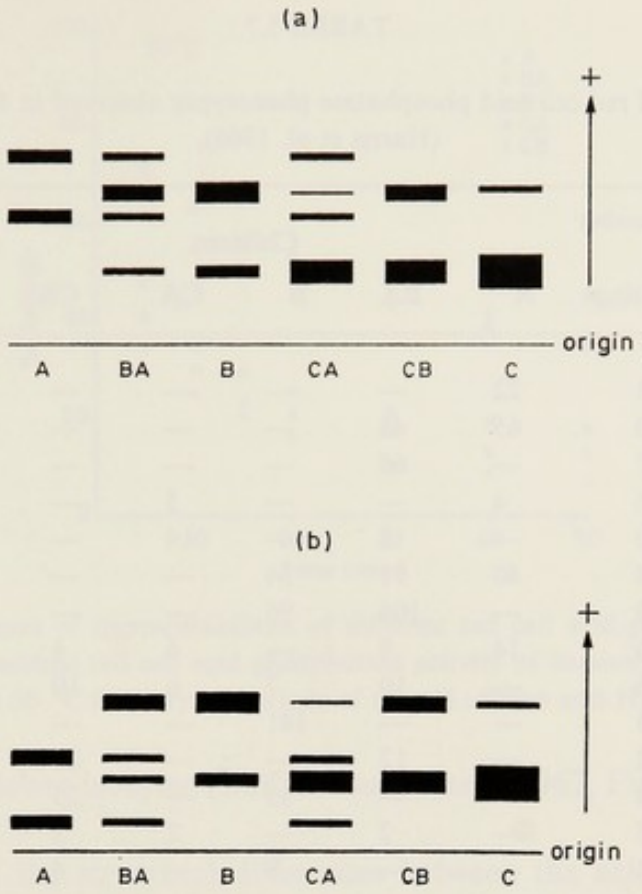


Fig. 5.10. Diagram showing electrophoretic patterns of different red cell acid phosphatase phenotypes. Starch gel electrophoresis at *p*H 6.0; (a) with citrate–phosphate buffer system; (b) with phosphate buffer system. Note that the two type A isozymes have a different relative mobility to the two type B and the two type C isozymes in the different buffer systems. For details of methods see Hopkinson and Harris (1969).

matings is most simply accounted for by the hypothesis that they are determined by three common autosomal alleles P^a, P^b and P^c. Types A, B and C represent the homozygous genotypes P^aP^a, P^bP^b and P^cP^c respectively, and types BA, CA and CB represent the heterozygous genotypes P^aP^b, P^aP^c and P^bP^c. Table 5.7 shows some typical family data and it will be seen that the findings are in good agreement with the Mendelian expectations.

The various electrophoretic phenotypes differ from one another in the electrophoretic mobilities, the relative activities and also in the number of isozymes present. The homozygous types A, B and C, each show two characteristic isozyme components. The two type A isozymes differ in their electrophoretic mobilities from each of the isozymes present in types B and types C. Also they are of roughly equal activities, whereas in types B and C there are marked disparities in the relative activities of the two isozymes

TABLE 5.7

The distribution of red cell acid phosphatase phenotypes observed in 440 English families (Harris et al. 1968).

Type of mating	Number of matings	Children						
		A	BA	B	CA	CB	C	Total
A×A	13	22	—	—	—	—	—	22
A×BA	50	65	48	—	—	—	—	113
A×B	27	—	65	—	—	—	—	65
A×CA	6	4	—	—	5	—	—	9
A×CB	10	—	5	—	14	—	—	19
BA×BA	94	40	91	54	—	—	—	185
BA×B	109	—	106	96	—	—	—	202
BA×CA	16	14	9	—	6	4	—	33
BA×CB	16	—	10	7	6	10	—	33
B×B	55	—	—	141	—	—	—	141
B×CA	12	—	12	—	—	20	—	32
B×CB	24	—	—	33	—	25	—	58
CA×CB	5	—	2	—	2	3	1	8
CB×CB	3	—	—	0	—	4	1	5
Totals	440	145	348	331	33	66	2	925

present. Thus in type B the faster, more anodic isozyme shows much more activity than the slower isozyme, whereas in type C the opposite is the case. The isozyme patterns seen in the heterozygous types BA, CA and CB are more complex and appear to represent simple mixtures in roughly equal proportions of the isozymes present in the two corresponding homozygous types. There is however no indication of hybrid isozyme formation in the heterozygotes.

Studies on the pattern of substrate specificity and on the kinetics of the acid phosphatases of the different types have not revealed any striking differences (Scott 1966, Luffman and Harris 1967). Also the various isozymes in the different types all appear to be of the same molecular size (Luffman and Harris 1967). However significant differences in thermostability have been demonstrated (fig. 5.11). In general the relative thermostabilities of the isozymes of the different types are in the order C > B > A, and within any type the faster (i.e. more anodic) isozyme appears to be somewhat less

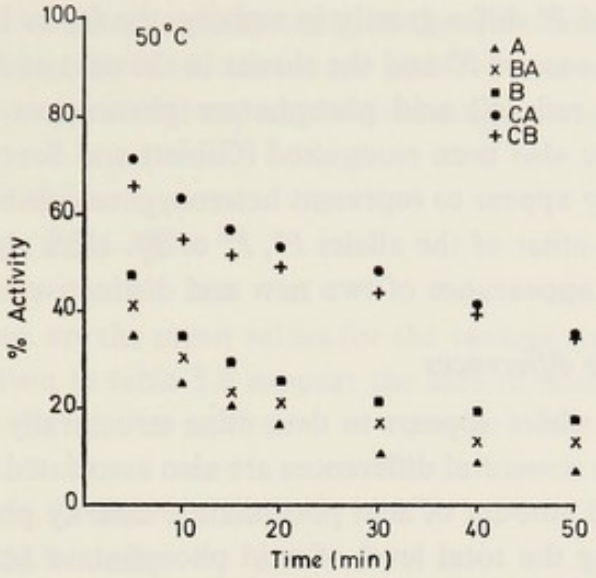


Fig. 5.11. Comparison of thermostabilities of different red cell acid phosphatase types. Mean percentage residual red cell acid phosphatase activity of haemolysates of different types held at 50 °C for varying periods of time (Luffman and Harris 1967).

stable than the slower isozyme (Luffman and Harris 1967, Fisher and Harris 1969).

The nature of the structural differences between the acid phosphatases determined by the several alleles is not known. One may suppose that the different alleles arose one from another by single mutational steps. Each presumably codes for a distinctive polypeptide chain, the differences between which may be no more than single aminoacid substitutions. Presumably also the two isozymic proteins which appear to be determined by each allele, both contain the same characteristic polypeptide chain, since they both seem to be modified in a characteristic way by the mutational differences. However the molecular basis for the two isozymes produced by the separate alleles is at present quite obscure. They appear to be of the same molecular size and there is no evidence that they contain more than one polypeptide chain, though this cannot be certainly excluded. They may represent conformational isomers or they may be the consequence of secondary modifications of the specific polypeptide coded by the particular allele, which are brought about subsequent to its primary synthesis. The allelic differences must lead not merely to the difference in molecular charge which seems to be the same for the pair of isozymes produced by each allele, but also to the striking differences in relative activity which are distinctive for each allele. Thus the two isozymic proteins from allele P^a show roughly equal activities, while those

from alleles P^b and P^c differ greatly in activity, the faster isozyme being the most active in the case of P^b and the slower in the case of P^c.

Besides the six red cell acid phosphatase phenotypes discussed above, several others have also been recognised (Giblett and Scott 1965, Karp and Sutton 1967). They appear to represent heterozygous combinations of a rare allele with one or other of the alleles P^a, P^b or P^c. Here also the new alleles each result in the appearance of two new and distinctive isozymic proteins.

5.4.2. Quantitative differences

Thus each of the alleles appears to determine structurally distinct forms of the enzyme. These structural differences are also associated with quantitative differences in total amount of acid phosphatase activity produced. This was shown by assaying the total level of acid phosphatase activity in red cells derived from individuals of the different types (Spencer et al. 1964).

Some typical findings are summarised in table 5.8. Although a great deal

TABLE 5.8

Average red cell acid phosphatase activity in individuals of different types. (Spencer et al 1964b.)

Type	Number of individuals tested	Mean activity	Standard deviation
A	33	122.4	16.8
BA	124	153.9	17.3
B	81	188.3	19.5
CA	11	183.8	19.8
CB	26	212.3	23.1

The activity is expressed as μmoles of p-nitrophenol liberated from p-nitrophenyl phosphate in 30 min at 37 °C, per gram of haemoglobin present in haemolysate.

of variation in the activity levels between individuals of the same acid phosphatase type occurs, nevertheless significant differences in average level of activity between the types are readily demonstrable. On average red cells of type B individuals show about 50% more activity than those of type A individuals, while type BA individuals have intermediate levels. Similarly the activity of red cells of type CB individuals is on average greater than those of type CA or type B.

Using this kind of data one may examine the question as to whether the quantitative effects of the three alleles are additive in a simple way or not. If they are additive one would expect the following relationships to be true:

$$\tfrac{1}{2}\overline{A} + \tfrac{1}{2}\overline{B} = \overline{BA} \qquad \text{(a)}$$

and $$\overline{CA} - \tfrac{1}{2}\overline{A} = \overline{CB} - \tfrac{1}{2}\overline{B} \qquad \text{(b)}$$

where $\overline{A}$, $\overline{BA}$, $\overline{B}$ etc. are the mean values for the various types. It will be seen that the results given in table 5.9 support the idea of additivity rather well. Thus

$$\tfrac{1}{2}\overline{A} + \tfrac{1}{2}\overline{B} = 155.35 \text{ units} \qquad \text{(a)}$$

$$\overline{BA} = 153.9 \text{ units}$$

$$\overline{CA} - \tfrac{1}{2}\overline{A} = 122.6 \text{ units} \qquad \text{(b)}$$

$$\overline{CB} - \tfrac{1}{2}\overline{B} = 118.15 \text{ units}$$

These results are of course consistent with the electrophoretic findings that the isozyme patterns in the heterozygous types are essentially those to be expected from simple mixtures in equal proportions of the two correspondingly homozygous types.

One point of some general significance which emerges from these studies is the following. If one determines red cell acid phosphatase activities in a series of randomly selected individuals in the general population one obtains a continuous unimodal distribution without any obvious discontinuities. It is not in fact dissimilar in form to the distributions often obtained when other enzymes are examined quantitatively in randomly selected populations. Yet in this case it is apparent that the overall distribution represents a summation of a series of separate but overlapping distributions corresponding to each of the discrete phenotypes (fig. 5.12). Furthermore much of the variance of the overall distribution can be attributed simply to the effects of the three alleles. This suggests that other examples of quantitative enzyme variation which are apparently continuous and unimodal may have a similar simple underlying basis. It also illustrates how difficult the genetical analysis of such quantitative enzyme variation may be in the absence of other methods for discriminating between the discrete phenotypes.

The precise way in which the structural differences between the acid phosphatases determined by the different alleles are related to these quantitative differences in activity remains to be elucidated. It is perhaps, at least in part, a consequence of the differences in enzyme stabilities, since the order

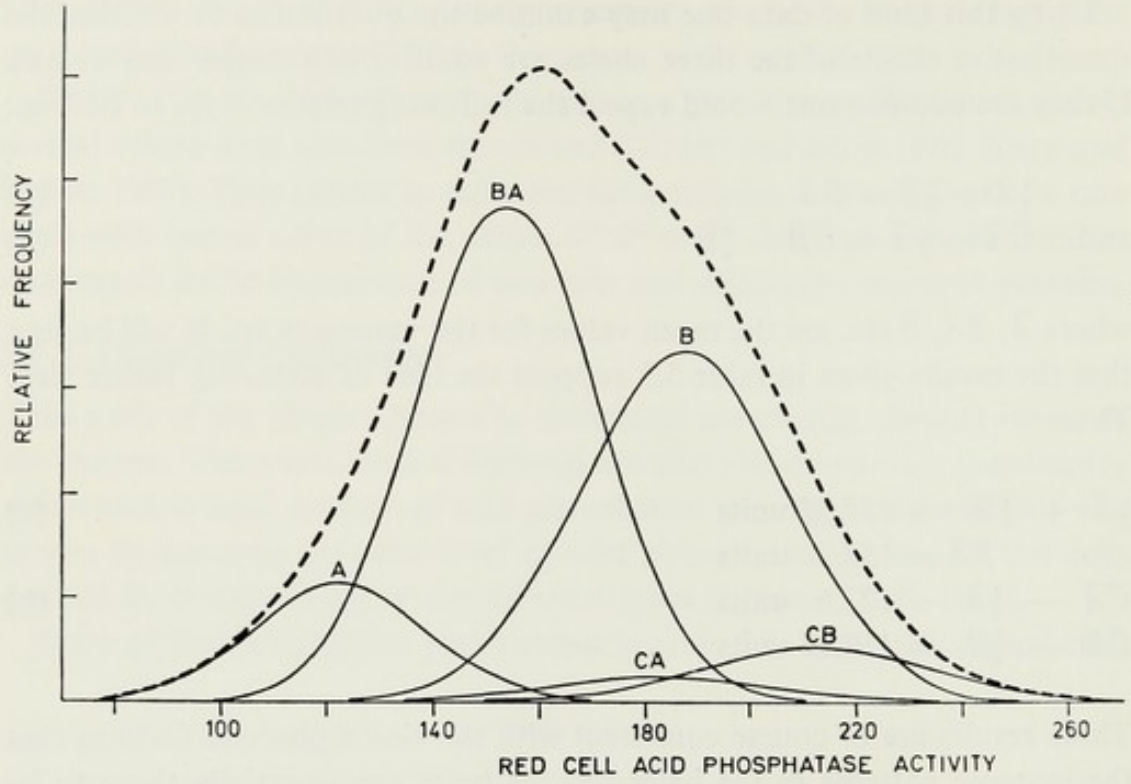


Fig. 5.12. Distribution of red cell phosphatase activities in the general population (broken line) and in the separate phenotypes. The curves are constructed from the data in table 5.8, using the frequencies of the different phenotypes as found in the English population.

of relative stabilities C > B > A found in thermostability experiments is the same as the order of relative activities in red cells of the different types.

CHAPTER 6

The inborn errors of metabolism

6.1. Garrod and the concept of 'inborn errors of metabolism'

It has been recognised for many years that there are a large number of metabolic diseases in which the characteristic clinical, pathological and biochemical abnormalities can be attributed to the congenital deficiency of a specific enzyme, which in turn is due to the presence of a particular abnormal gene. Such conditions are usually called 'inborn errors of metabolism', a term which was first used by A. E. Garrod more than sixty years ago.

Garrod's basic concept of the pathogenesis of these disorders (Garrod 1909) was mainly derived from studies on the rare condition known as alkaptonuria. His classical work on this abnormality has provided an elegant and simple model for the interpretation of a great variety of different inherited diseases subsequently discovered.

Alkaptonuria is characterised by the urinary excretion in massive quantities of the substance homogentisic acid, which is not normally found in urine. Alkaptonurics excrete several grams of homogentisic acid daily, and this is continuous and lifelong. It is a very striking abnormality because although the urine has a normal colour when passed, it rapidly goes black on standing due to the oxidation of the homogentisic acid present. The disorder is indeed often first recognised in infancy because of the characteristic staining of the diaper. Alkaptonurics are usually quite healthy though in later life they are particularly prone to develop a form of arthritis known as ochronosis, apparently caused by the deposition in cartilage and other connective tissue of a pigment derived from homogentisic acid.

Garrod noted that when homogentisic acid is fed to alkaptonuric subjects it is excreted quantitatively in the urine, whereas when given to normal subjects it appears to be readily metabolised. He also showed that its excretion in alkaptonurics could be augmented by feeding increased protein, and that this was due to the presence in protein of the aromatic aminoacids phenylalanine and tyrosine, which when given alone also enhanced the homogentisic acid

output. The excretion of homogentisic acid was also increased by certain derivatives of phenylalanine and tyrosine which could be plausibly regarded as intermediates in their catabolism.

From such metabolic experiments Garrod inferred that homogentisic acid, although it had never been detected in tissues, was itself a normal intermediate in the catabolism of phenylalanine and tyrosine, and that in alkaptonuric subjects the essential defect was a failure in its degradation, due to the lack of a necessary enzyme. Thus he supposed that in normal individuals homogentisic acid occurs at most only in trace amounts because it is broken down virtually as rapidly as it is formed. Whereas in alkaptonurics it cannot be broken down further, so it tends to accumulate in the cells of the liver where this metabolic process mainly occurs, leaks into the circulation and is excreted into the urine in large quantities.

At that time the crucial piece of evidence namely a direct demonstration of the specific enzyme deficiency was not available. In fact this was not obtained until some 50 years later when it became possible to assay in liver

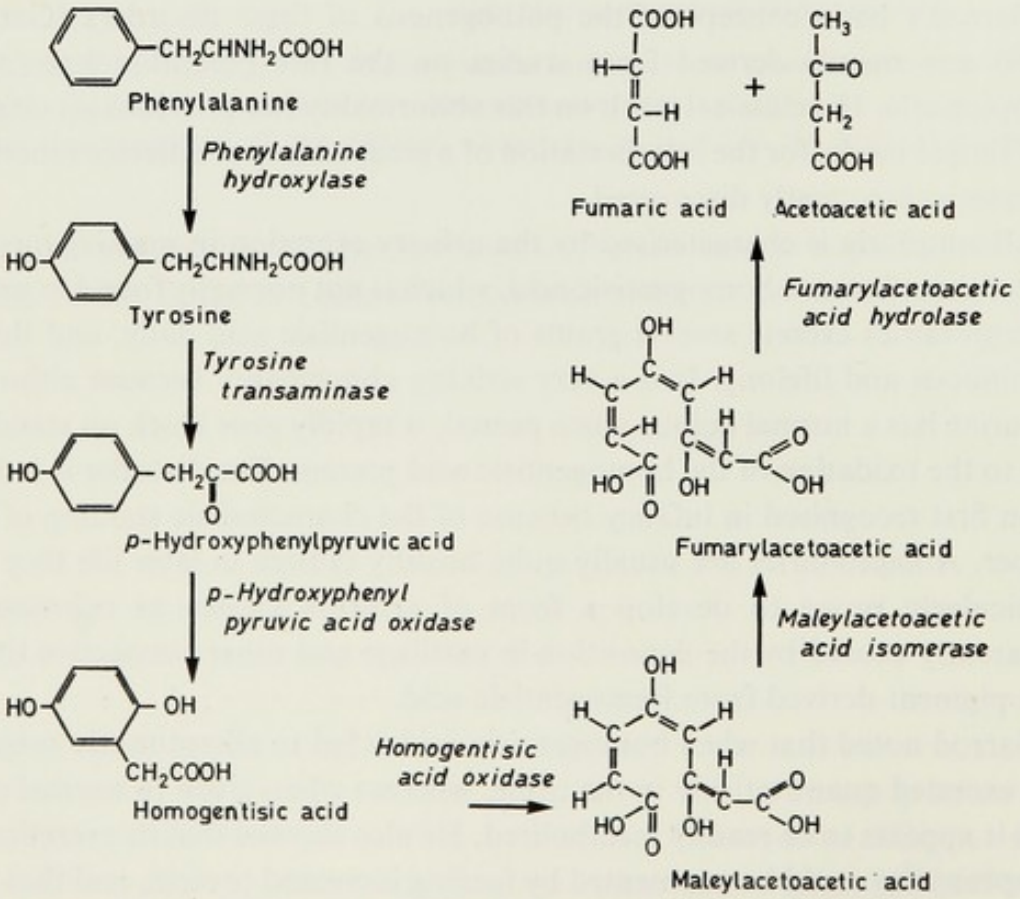


Fig. 6.1. Enzymatic steps in the oxidation of phenylalanine and tyrosine to acetoacetic acid. (From La Du 1966.)

biopsies from alkaptonuric subjects each of the series of enzymes concerned in the oxidation of tyrosine (fig. 6.1). It was found that all were present in normal amounts except for homogentisic acid oxidase, the enzyme which catalyses the conversion of homogentisic acid to maleylacetoacetic acid (La Du et al. 1958). This was not detectable (table 6.1). Thus Garrod's classical interpretation of the biochemistry of alkaptonuria was fully substantiated.

TABLE 6.1

Activity of tyrosine oxidation enzymes in biopsy specimens of alkaptonuric and non-alkaptonuric liver (La Du et al. 1958).

Enzyme	Enzyme activity	
	Non-alkaptonuric liver	Alkaptonuric liver
1. Tyrosine transaminase	3.6	3.2
2. p-Hydroxyphenylpyruvic acid oxidase	6.7	4.6
3. Homogentisic acid oxidase	26.8	<0.0048
4. Maleylacetoacetic acid isomerase	960	780
5. Fumarylacetoacetic acid hydrolase	29	22

Activity of enzymes 1, 2, 3 and 5 expressed as μmoles of substrate oxidized per hour by 0.1 g wet weight of liver. Activity of enzyme 4 expressed as Δ log O.D. per hour per 0.1 g wet weight of liver.

The other striking feature of alkaptonuria to which Garrod drew attention was its familial distribution. Although a very rare condition it was often found among more than one member of the family. Frequently two or more of a group of brothers and sisters would be affected, while their parents, children and other relatives appeared quite normal. Furthermore the parents of alkaptonurics were often blood relatives, usually first or second cousins. The pedigrees were quite characteristic, and Garrod had little hesitation in concluding that they implied an hereditary or genetical basis for the condition. He consulted Bateson, one of the earliest geneticists, who pointed out that the situation could be readily explained in terms of the then recently rediscovered laws of Mendel. The pedigrees were exactly those to be expected if alkaptonuria was determined by a rare recessive Mendelian factor, or as we should now say gene. The affected individuals could be presumed to be

homozygous for the abnormal gene. This was in fact the first example of the so-called 'recessive inheritance' to be recognised as such in man.

Thus Garrod interpreted alkaptonuria as being caused by the congenital deficiency of a particular enzyme due to the presence in double dose of an abnormal Mendelian factor or gene. An important implication of the idea was that the normal allele of this gene must in some way be necessary for the formation of the enzyme in the normal organism. This was the first clue to the now well established generalisation that genes exert their effects in the organism by directing the synthesis of enzymes and other proteins.

Garrod viewed the inborn errors as conditions in which the specific enzyme deficiency effectively blocked at a particular point a sequence of reactions which form part of the normal course of metabolism. As a result metabolites immediately preceding the block would accumulate, and metabolites subsequent to the block would not be formed. The various biochemical, pathological and clinical manifestations of the condition could be regarded as secondary consequences of this primary metabolic defect. These secondary changes might be complex and widespread, and would depend in general on the nature and the biochemical effects of the metabolites which tended to accumulate or whose formation was restricted.

A large number of different disorders which can be explained in these general terms are now known. They are listed in Appendix I, pp. 266–283. Although a specific enzyme deficiency has been demonstrated in each of these conditions, the underlying nature of the defect is understood in very few. In some cases it may represent the synthesis of a structurally altered enzyme protein which has defective catalytic properties. In other cases a structurally altered enzyme protein which is extremely unstable and so is rapidly broken down in the tissues, may be the cause of the enzyme deficiency. In still other cases there may be a specific reduction or complete failure in the synthesis of the enzyme protein.

The enzymes involved in these different disorders are very diverse and are concerned with many aspects of metabolism. The metabolic disturbances and clinical abnormalities which result also vary widely. They range from conditions which may be effectively lethal in early life (e.g. maple syrup urine disease), through those that produce a permanent disability such as mental retardation (e.g. phenylketonuria) or chronic haemolytic disease (e.g. pyruvate kinase deficiency), to those which are comparatively benign (e.g. alkaptonuria) or apparently harmless (e.g. fructokinase deficiency).

The examples discussed below illustrate something of the variety of

biochemical and metabolic disturbances that may follow from such specific enzyme deficiencies, and the clinical abnormalities that can ensue.

6.2. *Phenylketonuria*

Phenylketonuria is among the most common of the 'inborn errors' which give rise to severe clinical disability, and since its discovery by Fölling in 1934 it has been studied extensively. It is characteristically associated with a marked degree of mental retardation, and in most institutions for the mentally retarded about 0.5–1 % of the patients suffer from this disorder. In European populations the abnormality is probably present in about 1 in 15,000 newborn infants.

The deficient enzyme is phenylalanine 4-hydroxylase (Jervis 1953, Mitoma et al. 1957, Wallace et al. 1957) which in the normal individual occurs in the liver and catalyses para-hydroxylation of the aminoacid phenylalanine to give tyrosine. Phenylalanine is continuously being produced from the normal breakdown of tissue protein, and from the digestion of dietary protein (fig. 6.2). Its conversion to tyrosine in the liver is the first step in its catabolism and if this is blocked it accumulates intracellularly and appears in high

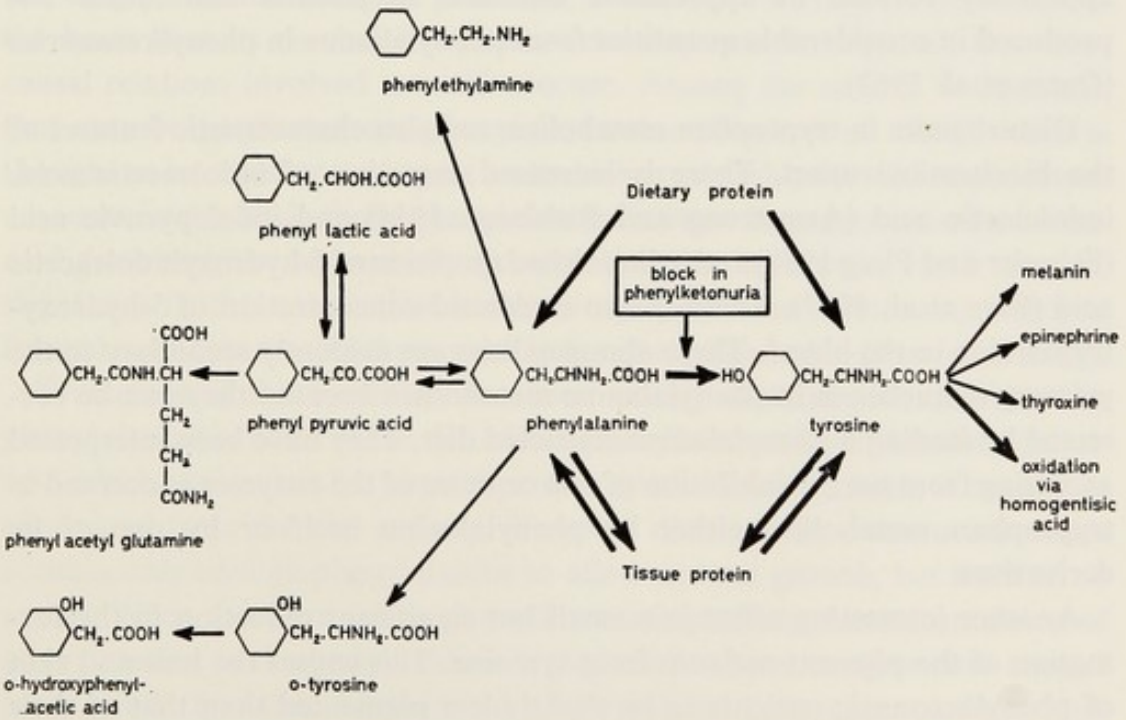


Fig. 6.2. Metabolic pathways involving phenylalanine, showing the site of the metabolic block in phenylketonuria.

concentrations in the body fluids. The level of phenylalanine in blood serum in phenylketonuria is generally more than thirty times normal, and there is an increased excretion of the aminoacid in urine. The level of phenylalanine in the cerebrospinal fluid is also considerably elevated.

The very high concentrations of phenylalanine that occur result in various secondary biochemical disturbances (fig. 6.2). One set of reactions involve changes in the side chain of phenylalanine (Jervis 1950). Large amounts of phenylpyruvic acid (Fölling 1934), phenyllactic acid (Zeller 1943) and phenylacetic acid are formed. The phenylacetic acid is subsequently conjugated with glutamine to give phenylacetyl glutamine (Woolf 1951). These substances have a low renal threshold and are excreted in the urine in considerable quantities. It was indeed the presence of phenylpyruvic acid in the urine that first led to the identification of phenylketonuria and gave it its name.

Another derivative of phenylalanine formed in abnormal amounts is o-hydroxyphenyl-acetic acid (Armstrong et al. 1955). Ortho-hydroxylation of phenylalanine (or phenylpyruvic acid) may occur in the normal individual (Tashian 1959) but is quantitatively insignificant compared with para-hydroxylation of phenylalanine to give tyrosine. When however the main pathway is blocked the product of the minor pathway is formed in increased amounts. Phenylethylamine is yet another derivative of phenylalanine not apparently formed in appreciable amounts in normal individuals but produced in considerable quantities from phenylalanine in phenylketonurics (Oates et al. 1963).

Disturbances in tryptophan metabolism are also characteristic features of the biochemical upset. There is increased excretion of indoleacetic acid, indolelactic acid (Armstrong and Robinson 1954) and indolepyruvic acid (Schreier and Flaig 1956) and a diminished excretion of 5-hydroxyindoleacetic acid (Pare et al. 1957). There is also a reduced concentration of 5-hydroxytryptamine in the blood. These abnormalities are evidently secondary to the primary disturbance in phenylalanine metabolism because they can be corrected by feeding a phenylalanine restricted diet. They have been interpreted as arising from partial inhibition of one or more of the enzymes concerned in tryptophan metabolism either by phenylalanine itself or by one of its derivatives.

Another interesting effect is a small but significant reduction in the formation of the pigment melanin from tyrosine. This causes the hair and skin of phenylketonuric patients to be slightly less pigmented than that of their normal sibs. It appears to be due to a partial inhibition of the enzyme tyrosinase by the high concentrations of phenylalanine which occur. Phenyla-

lanine acts as a competitive inhibitor of the tyrosine–tyrosinase system *in vitro* (Miyamoto and Fitzpatrick 1957), and it has been demonstrated that darkening of new grown hair in the phenylketonuric may be achieved either by increasing tyrosine intake considerably (Snyderman et al. 1955), or by severely restricting phenylalanine intake (Armstrong and Tyler 1955).

There is usually no tyrosine deficiency in phenylketonuria because adequate amounts of tyrosine are present in the diet. But it is of interest to note that while tyrosine is not an indispensible dietary ingredient (i.e. an 'essential' aminoacid) in the normal individual, it becomes one in phenylketonuria. This is because in the normal organism tyrosine is readily formed from phenylalanine, while in phenylketonuria this is not the case.

Physical development is not seriously affected by the metabolic disorder, but mental development is markedly impaired. The patients usually have a severe degree of intellectual retardation. The majority are graded as idiots (I.Q. less than 20) and the remainder nearly all as imbeciles (I.Q. less than 50), though some individuals with a lesser degree of impairment do occur.

The exact manner in which the damage to the brain is brought about is not understood. While it is reasonable to suppose that phenylalanine itself, or one of the other substances which are present in unusual concentrations in the body fluids may either by inhibiting certain enzyme systems, or by blocking particular transport processes, so modify the intracellular milieu in the brain that its normal biochemical development is impeded, the exact causal relations involved are still obscure. Among the specific compounds that may be implicated are phenylethylamine (Oates et al. 1963) which is thought to be a neurotoxic agent and which is probably formed in abnormal amounts in the brain from phenylalanine because the appropriate L-amino acid decarboxylase is normally present. Another substance which may be significant in this connection is 5-hydroxytryptamine (serotonin) which is formed in less than normal amounts in the phenylketonuric (Pare et al. 1957).

Treatment of the condition has been mainly directed to restricting the phenylalanine content of the diet. Since phenylalanine is an essential amino-acid which is required for normal protein synthesis and growth, it cannot be eliminated from the diet entirely. However diets can be constructed which contain only enough phenylalanine to allow normal growth, but little or no excess. When these are fed to phenylketonuric patients the concentration of phenylalanine in the body fluids is brought down to normal or near normal levels, and the other biochemical abnormalities consequent on the high phenylalanine concentrations disappear (Armstrong and Tyler 1955). To what degree the continuous administration of such diet will minimise or

prevent mental impairment is still uncertain, because of the very extensive and long term studies which are needed to assess the value of the treatment objectively. But the results so far are not unhopeful. It has however become clear that if any degree of success is to be achieved with this line of therapy, it is important that the diet should be started as early as possible, since most of the brain damage in phenylketonuria probably occurs in the first few months after birth.

6.3. Galactosaemia

Galactosaemia is an inborn error of carbohydrate metabolism in which there is a specific inability to metabolise the hexose sugar galactose (Townsend et al. 1951, Komrower et al. 1956). Galactose is an important constituent of an infant's diet because it is a component of the disaccharide lactose which is the main carbohydrate present in milk. Thus an infant affected with this disorder who is fed on milk in the ordinary way will receive considerable quantities of galactose. The consequences are generally severe. The infant fails to thrive, weight gain is slow, mental development is retarded, the liver becomes enlarged and eventually cirrhotic, and cataracts develop in the lens. Death in infancy is not uncommon if the condition is not recognised. If however such an infant is placed on a diet entirely free of galactose a dramatic improvement in the physical condition takes place. In fact it seems probable that provided the treatment is started early enough and great care is taken rigorously to exclude galactose from the food, growth and development may take place normally. If however commencement of treatment is delayed, then some degree of liver damage, cataract and mental impairment may persist because irreversible changes have already taken place.

Galactose enters the main stream of carbohydrate metabolism via a series of reactions which result in its conversion to glucose-1-phosphate (fig. 6.3). It first reacts with ATP to give galactose-1-phosphate. The enzyme concerned is galactokinase. The galactose-1-phosphate then reacts with the nucleotide uridine diphosphoglucose to give glucose-1-phosphate and uridine diphosphogalactose. The enzyme here is galactose-1-phosphate uridyl transferase. Uridine diphosphoglucose can then be regenerated from uridine diphosphogalactose by the enzyme uridine diphosphogalactose-4-epimerase, a reaction which requires the coenzyme NAD. In galactosaemia the enzyme galactose-1-phosphate uridyl transferase is virtually absent (Kalckar et al. 1956, Isselbacher et al. 1956). Galactokinase and the epimerase occur in normal amounts as do other enzymes concerned in carbohydrate metabolism.

When galactose is present in the diet, the enzyme deficiency results in an intracellular accumulation of galactose-1-phosphate (Schwarz et al. 1956) and an abnormally high level of galactose itself in the body fluids. The blood galactose is considerably elevated after a galactose containing meal, and declines only slowly. Because of the high blood levels, large amounts of galactose appear in the urine. The elevated galactose concentrations in the body fluids also lead to the formation of the sugar alcohol galactitol in abnormal amounts (Wells et al. 1964, 1965).

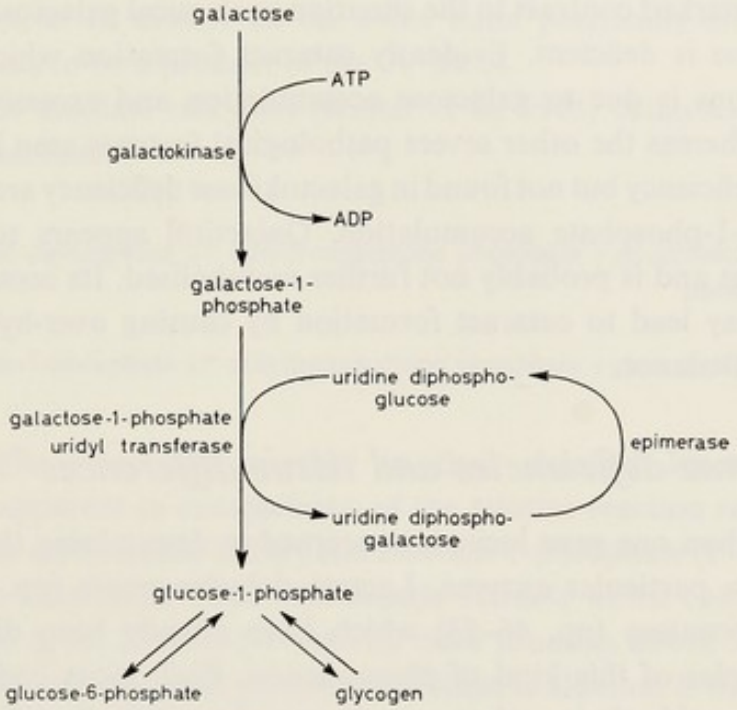


Fig. 6.3. Pathways in galactose metabolism.

Most of the pathological consequences that occur in galactosaemia can be attributed to the high intracellular concentrations of galactose-1-phosphate that develop. This is thought to produce pathological effects by inhibiting other enzyme reactions in carbohydrate metabolism which involve phosphorylated intermediates. The liver and brain damage as well as the general failure to thrive are probably secondary consequences of this. A notable feature is a tendency to hypoglucosaemia due to a decreased hepatic output of glucose.

The development of cataracts of the lens, which is a typical feature of the disease, is however probably due to the high concentrations of galactose in the body fluids and the resultant formation of abnormal quantities of galactitol (Gitzelmann et al. 1967). In this connection it is interesting to

compare the pathological changes which occur in galactosaemia due to galactose-1-phosphate uridyl transferase deficiency with those found in another inborn error of galactose metabolism, where there is a gross deficiency of galactokinase, but normal levels of the uridyl transferase and the epimerase. In galactokinase deficiency (Gitzelmann 1967), high blood levels of galactose occur when milk is fed, and galactitol is produced in abnormal amounts. Galactose-1-phosphate however is not formed. The outstanding clinical feature is the development of severe lens cataracts from an early age. In other respects the affected individuals may be healthy and develop normally. This is in marked contrast to the situation in classical galactosaemia where the transferase is deficient. Evidently cataract formation which occurs in both conditions is due to galactose accumulation and excessive galactitol formation, whereas the other severe pathological features seen in the uridyl transferase deficiency but not found in galactokinase deficiency are consequent on galactose-1-phosphate accumulation. Galactitol appears to be readily formed in lens and is probably not further metabolised. Its accumulation in this tissue may lead to cataract formation by causing over-hydration and electrolyte unbalance.

6.4. Isozyme deficiencies and tissue differences

Often more than one gene locus is concerned in determining the molecular structure of a particular enzyme. Lactate dehydrogenase (pp. 40–46) and phosphoglucomutase (pp. 46–53) which have already been discussed are typical examples of this kind of phenomenon. Each locus codes for a distinctive polypeptide chain and as a result several structurally different molecular forms or isoenzymes of the enzyme occur. Also it is not infrequently found that the relative amounts of the different isoenzymes vary considerably from tissue to tissue, presumably because of variations in expression of the several gene loci in different cell types.

In such cases, a mutation at one of the loci may result in a deficiency of some isozymes but not of others. Furthermore if tissues or organs differ in the isozymes they normally contain, then this will be reflected in the biochemical and clinical manifestations of a particular mutant. The characteristic and sometimes rather unexpected features of many inherited diseases are often explicable in terms of such isozymic differences in the normal organism.

6.4.1. Aldolase deficiency in hereditary fructose intolerance

At least three structurally distinct forms of the enzyme aldolase have been

shown to occur (Penhoet et al. 1966, 1967). These are known as aldolase A which is the form found in muscle, aldolase B which is the predominant form in liver, and aldolase C which occurs in brain. These isozymic proteins are tetramers and evidently differ in the structures of their characteristic polypeptide subunits, each of which is presumably determined by a separate gene locus. The tissue differences seem to reflect differences in the relative amounts of synthesis of the polypeptide products of the different loci. Thus in liver the polypeptide determined by the 'B' locus is formed in much greater amounts than that determined by the 'A' locus, and the 'C' locus is probably virtually inactive. In muscle on the other hand practically all the enzyme formed appears to be a product of the 'A' locus.

Each of the aldolase isozymes (Rutter et al. 1968) catalyses both of the following reactions:

(1) Fructose diphosphate $\rightleftarrows$ dihydroxyacetone phosphate + glyceraldehyde-3-phosphate

(2) Fructose-1-phosphate $\rightleftarrows$ dihydroxyacetone phosphate + glyceraldehyde.

But they differ from one another in their detailed kinetics. This is particularly apparent in comparisons of the relative reaction rates obtained when fructose diphosphate (FDP) and fructose-1-phosphate (F-1-P) are used separately as substrates. Thus with muscle extracts which contain virtually only aldolase A the FDP:F-1-P activity ratio found is about 50:1, whereas with liver extracts where the predominant form is aldolase B the FDP:F-1-P activity ratio is about 1:1.

Hereditary fructose intolerance (Chambers and Pratt 1956, Froesch et al. 1957) appears to be specifically an inborn error of fructose metabolism. Individuals with the abnormality, whether they are in infant or adult life, remain healthy and symptom free as long as they do not take any food containing fructose. The ingestion of fructose however, whether as the free form or in the disaccharide sucrose, causes immediate deleterious effects the main symptoms being attributable to a severe hypoglucosaemia which develops. The condition is usually first recognised in infancy, since the symptoms are likely to become manifest as soon as breast feeding is supplemented or replaced by feeds which contain added sucrose. There is often a rapid deterioration in the child's condition, and the disorder may prove fatal if it is not recognised and sucrose promptly removed from the diet (fig. 6.4).

Fructose enters the main stream of carbohydrate metabolism, by first

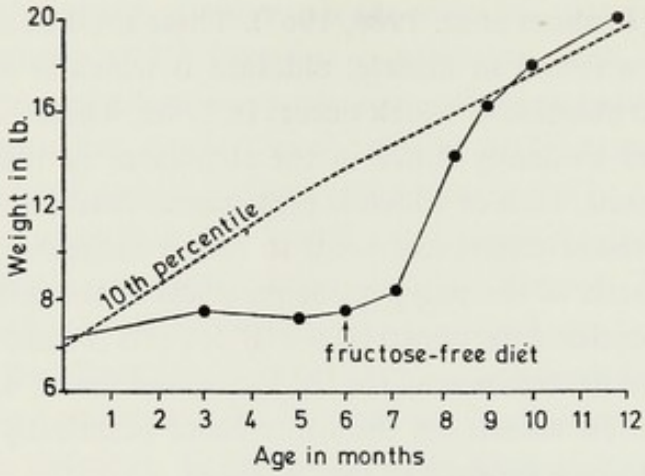


Fig. 6.4. Weight chart of a child with fructose intolerance. He was breast fed for only two days and then put on a dried milk preparation with added sucrose. There was virtually no weight gain over the next few months, and his condition deteriorated. He only began to improve and gain weight when a fructose-free diet was instituted (Black and Simpson 1967).

being converted to fructose-1-phosphate with ATP and the enzyme fructokinase (fig. 6.5). The fructose-1-phosphate is then cleaved by aldolase

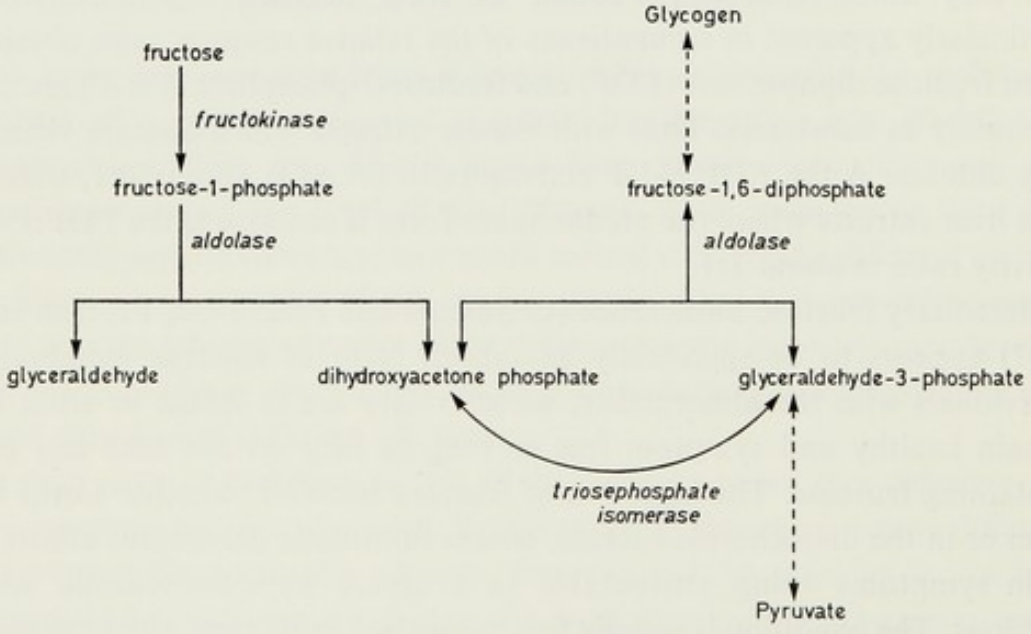


Fig. 6.5. Pathways in fructose metabolism.

(reaction 2 above). In hereditary fructose intolerance a considerable reduction in aldolase activity in liver but not in muscle has been demonstrated (Hers and Joassin 1961, Froesch et al. 1963). Furthermore the reduction of activity

in liver is much more pronounced when fructose-1-phosphate is used as substrate for the enzyme assay, than when fructose diphosphate is used. Thus in one study (table 6.2) it was found that with fructose-1-phosphate as

TABLE 6.2

Aldolase activity in liver biopsy specimens from patients with fructose intolerance and from appropriate controls (Hers & Joassin 1961).

Subjects		Liver aldolase activity (μmoles substrate utilised per min per g tissue)		Ratio of activities
		With fructose diphosphate (FDP)	With fructose-1-phosphate (F-1-P)	FDP:F-1-P
Patients with fructose intolerance	1	3.1	0.5	6.2:1
	2	2.5	0.4	6.2:1
	mean	2.8	0.45	6.2:1
Controls	1	7.2	8.0	0.90:1
	2	15.4	15.8	0.97:1
	3	8.5	9.1	0.93:1
	4	11.0	12.4	0.89:1
	5	14.9	13.7	1.09:1
	mean	11.4	11.8	0.96:1

substrate the aldolase activity in liver specimens from the affected patients was only about 4% of that found in comparable controls. With fructose diphosphate as substrate the liver aldolase activity was about 25% of the control levels. The FDP:F-1-P activity ratios for the patients were around 6:1 compared with control values of about 1:1. Similar results have been obtained in other studies, and it is clear that a consistent feature of the disease is a marked reduction in liver aldolase F-1-P activity, a moderate reduction in liver FDP activity, and a considerable rise in the FDP:F-1-P activity ratio. In contrast to this, studies on muscle aldolase show no abnormality. Activity levels with both FDP and F-1-P are similar to those in the controls and the FDP-F-1-P ratio is unaltered.

These findings are most simply accounted for by a specific deficiency of

aldolase B. The residual aldolase activity in liver is probably mainly due to aldolase A, the characteristic polypeptide subunit of which is normally formed in liver but only in small quantities. It is also possible that there may be some compensating synthesis of aldolase A as a response to aldolase B deficiency. In muscle, where only the aldolase A polypeptide is normally formed, no aldolase deficiency occurs.

When fructose is administered, the aldolase B deficiency results in an intracellular accumulation of fructose-1-phosphate and an abnormal elevation of fructose in the body fluids. The blood fructose levels are high and fructose is excreted in the urine. The toxic effects of fructose in these circumstances are almost certainly attributable to the grossly increased concentrations of fructose-1-phosphate in the cells of the liver, because in another abnormality of fructose metabolism due to fructokinase deficiency (Schapira et al. 1961) no clinical disorder occurs at all, even though large amounts of fructose are given. Here high blood fructose levels develop as in aldolase deficiency, but there is no accumulation of fructose-1-phosphate. The precise manner in which fructose-1-phosphate exerts its toxic effects in hereditary fructose intolerance is not known, but it seems likely that it does so by inhibiting other enzymes involved in carbohydrate metabolism in liver, particularly those concerned in glycogen breakdown and the maintenance of a normal blood sugar level.

In effect the specific aldolase abnormality in hereditary fructose intolerance results only in defective metabolism of ingested fructose. Glycolysis in muscle is not affected because aldolase A activity is normal. Glycolysis and gluconeogenesis in liver are not seriously disturbed because the enzyme activity necessary for fructose diphosphate cleavage or synthesis is sufficient for normal requirements provided fructose is omitted from the diet and therefore no secondary disturbances due to fructose-1-phosphate accumulation occur.

6.4.2. Pyruvate kinase deficiency

Pyruvate kinase catalyses the conversion of phosphoenolpyruvate to pyruvate, a key step in glycolysis which is coupled to the generation of ATP:

$$\text{2-phosphoenolpyruvate} + \text{ADP} \rightleftarrows \text{pyruvate} + \text{ATP}.$$

At least two distinct isozymic forms of pyruvate kinase have been shown to occur (Bigley et al. 1968). They differ in their kinetic behaviour, as well as in various physical properties, and they can also be distinguished immuno-

chemically. They are presumably determined by separate gene loci. In man one of these forms has been found only in red cells and in liver. The other is more widely distributed and has been found in liver, kidney, skeletal muscle, cardiac muscle and leucocytes, but it is not present in red cells.

Many examples of a form of chronic haemolytic anaemia, apparently due to a specific deficiency of red cell pyruvate kinase, have been identified (Valentine et al. 1961, Bowman and Procopio 1963, Grimes et al. 1964, Keitt 1966, Tanaka and Valentine 1968). The enzyme deficiency causes a major disturbance of red cell glycolysis with impairment of energy supplies and shortening of the average life of the red cells. The degree of enzyme deficiency is usually severe, though it varies somewhat between patients from different families, and it appears that it may be brought about by a number of different mutant alleles. In some cases the defective enzyme has been shown to have altered kinetics (Boivin and Galand 1967, Paglia et al. 1968).

Even though the level of pyruvate kinase in the red cells is grossly reduced in these cases, the level of the enzyme in leucocytes is quite normal. So the defect is evidently peculiar to the red cell type of isozyme. The other tissue in the body in which the red cell type of isozyme occurs is liver, but here it is present together with the isozymic form also found in leucocytes. Thus one would expect to observe some degree of deficiency of liver pyruvate kinase activity in patients with the red cell deficiency. This has indeed been demonstrated (Bigley and Koler 1968), and it was also shown that the reduction in the total level of activity in the liver was due to the reduction of the isozymic form which also occurs in the red cell, whereas the other isozyme was unaffected. Evidently in such circumstances the amount of activity contributed by the unaffected isozyme is adequate to maintain normal function, so that no significant metabolic disturbance in the liver is apparent.

6.5. Partial enzyme deficiencies and their metabolic consequences. The urea cycle enzymes

The classical idea of an inborn error of metabolism was a situation where the specific enzyme concerned was absent so that the metabolic pathway it subserved was completely blocked. Metabolites immediately preceding the block would be expected to accumulate while the normal products of the reaction would not be formed. As more and more examples have been studied it has become clear that in many inherited metabolic diseases although there is certainly a specific enzyme deficiency, the loss of activity is not complete. Consequently some formation of the normal reaction products goes on, and

the metabolic pathway is therefore only partially blocked. Nevertheless the restriction on the formation of the products of the reaction as well as the accumulation of metabolites immediately preceding the partial block are often sufficient to give rise either directly or indirectly to serious clinical abnormalities. In other cases of course although some distortion of the normal metabolic situation may be evident, no clinical abnormality ensues except perhaps under special conditions of metabolic stress.

According to the degree of the enzyme deficiency and the particular nature of the metabolic pathway involved a wide range of different effects may occur. One particular type of situation is of special interest because the biochemical findings may at first sight seem somewhat anomalous. This is where the product of a reaction or reaction sequence appears to be formed at normal or near normal rates, but where the occurrence of grossly increased concentrations of certain intermediary metabolites in the reaction sequence indicates that a quite severe metabolic block is nevertheless present, and indeed a marked though partial deficiency of the appropriate enzyme may be demonstrable. This type of situation can often be explained in terms of an increased reaction velocity produced by elevation in substrate concentration. Although the specific enzyme may be much reduced compared with the normal, the increased concentration of its substrate which this reduction inevitably produces may be sufficient to raise the reaction velocity, that is the rate of formation of the product, to normal or near normal levels.

The general point is illustrated by the biochemical findings (Efron 1966) in the series of rare disorders involving specific deficiencies of enzymes in the Krebs–Henseleit cycle of reactions (fig. 6.6). This is the well-known reaction sequence by which urea, the principal nitrogenous excretory product of the organism, is produced.

One condition is called argininosuccinicaciduria, and is due to a deficiency of the enzyme argininosuccinase (argininosuccinate arginine lyase) which cleaves argininosuccinate to give arginine and fumarate (Allan et al. 1958, Westall 1960). Argininosuccinate, an intermediate in the urea cycle and not normally present in more than trace amounts, appears in considerable quantities. Its level in serum is elevated and it is excreted in the urine in large amounts. There is also some increase in the serum level of citrulline, its immediate precursor in the cycle. Blood ammonia levels, which may be normal in the fasting state, tend to rise significantly after meals containing any substantial amount of protein. But the blood urea level is not reduced below normal levels.

Urea is mainly formed in liver, and a marked and quite specific deficiency

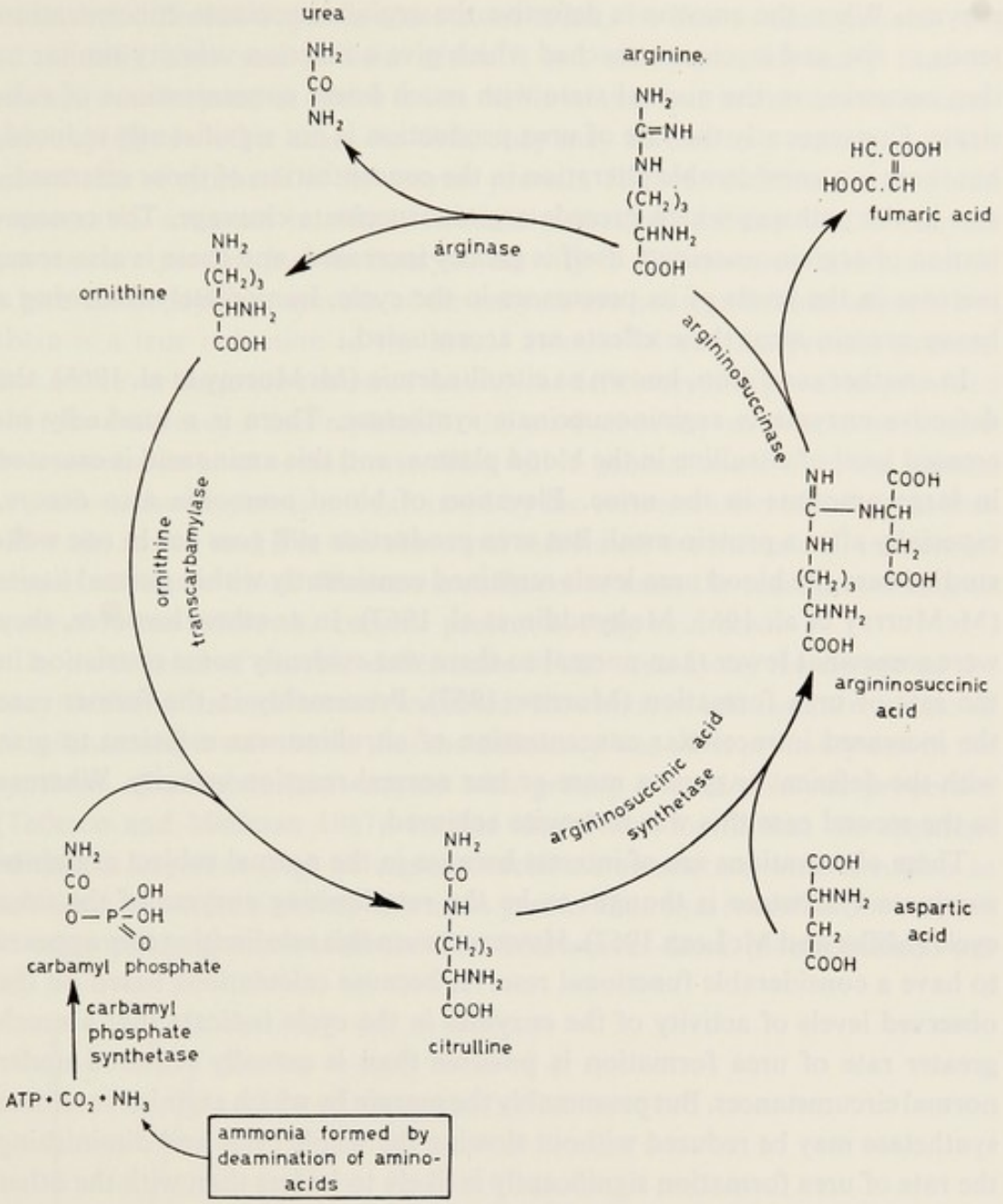


Fig. 6.6. Enzymes of the urea cycle.

of argininosuccinase activity in liver samples from patients with this condition has been demonstrated. But although less than 5% of the normal activity of the enzyme (Miller and McLean 1967) is apparently present, there is evidently still sufficient activity to allow urea synthesis to proceed at a more or less normal rate. Presumably, in the normal, the concentration of the intermediate metabolite argininosuccinate is very low and far from saturating the

enzyme. When the enzyme is defective the argininosuccinate concentration tends to rise and levels are reached which give a reaction velocity similar to that occurring in the normal state with much lower concentrations of substrate. Consequently the rate of urea production is not significantly reduced, but there is a considerable alteration in the concentration of those intermediates in the pathway which precede argininosuccinate cleavage. The concentration of argininosuccinate itself is greatly increased, and there is also some increase in the levels of its precursors in the cycle. Immediately following a heavy protein meal these effects are accentuated.

In another condition, known as citrullinaemia (McMurray et al. 1963), the defective enzyme is argininosuccinate synthetase. There is a markedly increased level of citrulline in the blood plasma, and this aminoacid is excreted in large amounts in the urine. Elevation of blood ammonia also occurs, especially after a protein meal. But urea production still goes on. In one well-studied case the blood urea levels remained consistently within normal limits (McMurray et al. 1963, Mohyuddin et al. 1967). In another, however, they were somewhat lower than normal so there was evidently some restriction in the rate of urea formation (Morrow 1967). Presumably in the former case the increased intracellular concentration of citrulline was sufficient to give with the deficient enzyme a more or less normal reaction velocity. Whereas in the second case this was not quite achieved.

These observations are of interest because in the normal subject argininosuccinate synthetase is thought to be the rate limiting enzyme of the urea cycle (Miller and McLean 1967). However, even this rate limiting step appears to have a considerable functional reserve, because calculations based on the observed levels of activity of the enzymes in the cycle indicate that a much greater rate of urea formation is possible than is actually required under normal circumstances. But presumably the margin by which argininosuccinate synthetase may be reduced without slowing the whole cycle and diminishing the rate of urea formation significantly is likely to be less than with the other enzymes of the cycle. It may be reached in certain cases of citrullinaemia and not in others.

In yet another condition known as 'hyperammonaemia' a deficiency of ornithine transcarbamylase has been reported (Russell et al. 1962). This enzyme catalyses the formation of citrulline from ornithine and carbamyl phosphate (fig. 6.6). Here again the enzyme deficiency is incomplete and the level of blood urea is within normal limits. But there is a consistent elevation of blood ammonia.

As has been noted previously, a deficiency of a particular enzyme may arise

in several different ways. In some cases it may be due to the synthesis of a structurally abnormal protein which is less efficient catalytically than its normal counterpart, or which is less stable so that it is broken down excessively rapidly. In other cases the deficiency may be due to a reduction in the actual rate of synthesis of the enzyme protein. Thus the deficiency state could represent a situation in which essentially normal amounts of the enzyme protein are present but its catalytic activity is defective; or a situation in which the catalytic properties of the enzyme are not significantly altered, but there is a true reduction in the actual amount of enzyme protein present. But whatever the precise cause of the enzyme deficiency, it is possible, provided that the deficiency is incomplete, for the product of the reaction to be formed at a normal or near normal rate. This will in general occur if, in the normal organism, the enzyme is not saturated by the concentration of substrate that normally exists, and if in the abnormal individual the increase in concentration of substrate which develops is sufficient to raise the reaction velocity to normal levels before the enzyme present is fully saturated.

In most cases of enzyme deficiencies in the urea cycle, the nature of the enzyme defect has not been determined. However, in one instance, a case of citrullinaemia, it was shown that the deficiency was most probably due to the synthesis of an abnormal enzyme with grossly altered catalytic properties (Tedesco and Mellman 1967). In this case it was found that the apparent Michaelis constant (K_m) of argininosuccinate synthetase with citrulline as substrate was at least twenty-five times greater than that found for the enzyme in appropriate normal controls. The Lineweaver–Burk plots obtained are

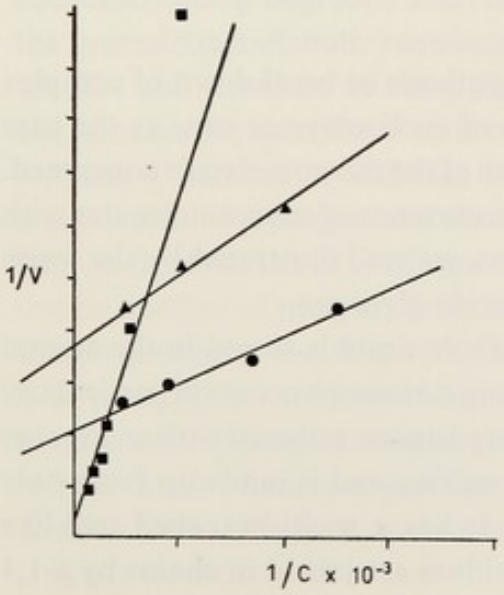


Fig. 6.7. Lineweaver–Burk plots derived from measurements of reaction velocities (V) of argininosuccinate synthetase at various concentrations (C) of citrulline (Tedesco & Mellman 1967). The enzyme was studied in extracts of fibroblast cells grown in tissue culture and derived from a patient with citrullinaemia, and from appropriate controls. *Key:* ■: enzyme from cells of citrullinaemic patient; ▲ and ●: enzymes from cells of two different control subjects.

shown in fig. 6.7. It will be seen that provided a sufficiently high intracellular concentration could be reached, the velocity of the reaction with the abnormal enzyme might be essentially the same as in the normal individual where the intracellular concentrations of citrulline are very low. In fact in this particular case this degree of compensation was probably not quite achieved *in vivo*, since the blood urea levels of the patient were somewhat less than normal. This was perhaps because the very considerable intracellular concentration of citrulline in the liver cells that would have been required to obtain a normal rate of urea formation could not be reached because of leakage of the citrulline from the liver cells into the body fluids. However, one can see that in other cases, if the elevation in K_m were not quite so severe, an essentially normal rate of urea formation might well occur.

In these various disorders of the urea cycle, the primary metabolic disturbance is in the liver, but the principal clinical abnormalities that ensue are largely attributable to damage to the central nervous system. A severe degree of mental retardation is a common feature. It appears probable that the neurological damage is mainly caused by the toxic effects of elevated blood ammonia levels, which tend to occur in each of these conditions particularly after heavy protein meals. This may be particularly important in infancy and early childhood when rapid maturation of the brain is taking place. So though in most of the cases it appears that the overall rate of production of urea is not markedly altered, severe clinical abnormality occurs because of the secondary effects of the accumulation of metabolites prior to the partial block in the sequence of reactions.

6.6. *Glycogen diseases*

Certain enzymes are concerned with the synthesis or breakdown of complex macromolecules, and specific deficiencies of such enzymes can, as the case may be, result in a deficit or an accumulation of the macromolecule concerned. They may also in some cases lead to the occurrence of macromolecules with a quite unusual structure. Such phenomena are well illustrated by the series of inborn errors involving the polysaccharide glycogen.

Glycogen is the main form in which carbohydrate is stored in the animal body. Although it is present in the cells of most tissues it occurs in particularly large amounts in liver and muscle. It is a polydisperse polymer with an average molecular weight of between 2.5 and 4.5 million and is made up from only one type of building block, α-D-glucose. It has a multi-branched tree-like structure (fig. 6.8). Most of the glucose residues are joined in chains by α-1,4

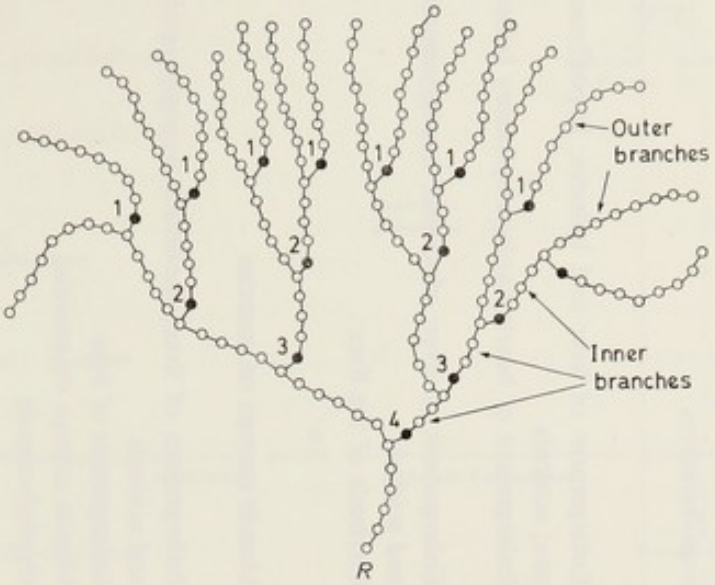


Fig. 6.8. Model of a segment of a glycogen molecule (after Cori 1954). There are 209 glucose residues, mol. wt. 33,858. Open circles: glucose residues in α-1,4 linkage; black circles: glucose residues in α-1,6 linkage. R = reducing end group. There are four tiers of branch points (glycogen has at least seven). Inner chains are terminated by branch points in adjacent tiers; outer chains by a branch point and by the non-reducing terminal glucose residue.

linkages. But at the branch points α-1,6 linkages occur. The outer chains which end in non-reducing terminal glucose residues tend to be longer than the inner chains which occur between two branch points. These outer chains usually consist of 7 to 10 glucose units and may constitute approximately 50% of the macromolecule. In normal individuals glycogen molecules are constantly being degraded and resynthesised to various degrees according to the immediate metabolic requirements. Thus the amount and to some extent the size and structure of the glycogen molecules present will depend at any one time on the nutritional state of the individual.

Quite a number of different enzymes are involved in the synthesis and degradation of glycogen and the main pathways are indicated diagrammatically in fig. 6.9. A variety of rare disorders each due to a specific deficiency of one or another of these enzymes have been recognised (table 6.3). In some of these conditions the defect has been found to involve reactions concerned with the synthesis of the macromolecule. But more often the defect appears to involve some particular step in glycogen degradation and results in the intracellular accumulation of glycogen in various tissues thus giving rise to what is often referred to as a 'glycogen storage disease'.

TABLE 6.3

Glycogen diseases (for references see text).

	Eponym	Enzyme defect	Glycogen structure	Main clinical manifestations
1.	von Gierke's disease (Cori type I)	Glucose-6-phosphatase	Normal	Enlargement of liver and kidneys; hypoglycaemia; acidosis
2.	Pompe's disease (Cori type II)	α-1, 4-glucosidase (lysosomal)	Normal	Enlargement of heart; cardio-respiratory failure
3.	Forbe's disease (Cori type III)	Amylo-1, 6- glucosidase (debrancher enzyme)	Abnormal; very short outer chains	Enlargement of liver; moderate hypoglycaemia and acidosis
4.	Andersen's disease; amylopectinosis (Cori type IV)	Amylo-(1,4 $\rightarrow$ 1,6)-transglucosidase (brancher enzyme)	Abnormal; long inner and outer chains with very few branch points	Cirrhosis of the liver
5.	McArdle's disease (Cori type V)	Muscle phosphorylase	Normal	Muscle cramps on exercise
6.	Hers' disease (Cori type VI)	Liver phosphorylase	Normal	Enlargement of liver; moderate hypoglycaemia and acidosis
7.	—	Phosphorylase kinase	Normal	Enlargement of liver
8.	—	Phosphofructokinase	Normal	Muscle cramps on exercise
9.	—	Glycogen synthetase (UDPG-glycogen-1, 4-glucosyltransferase)	—	Hypoglycaemia

'Cori types' based on classification proposed by G. Cori (1954, 1957).

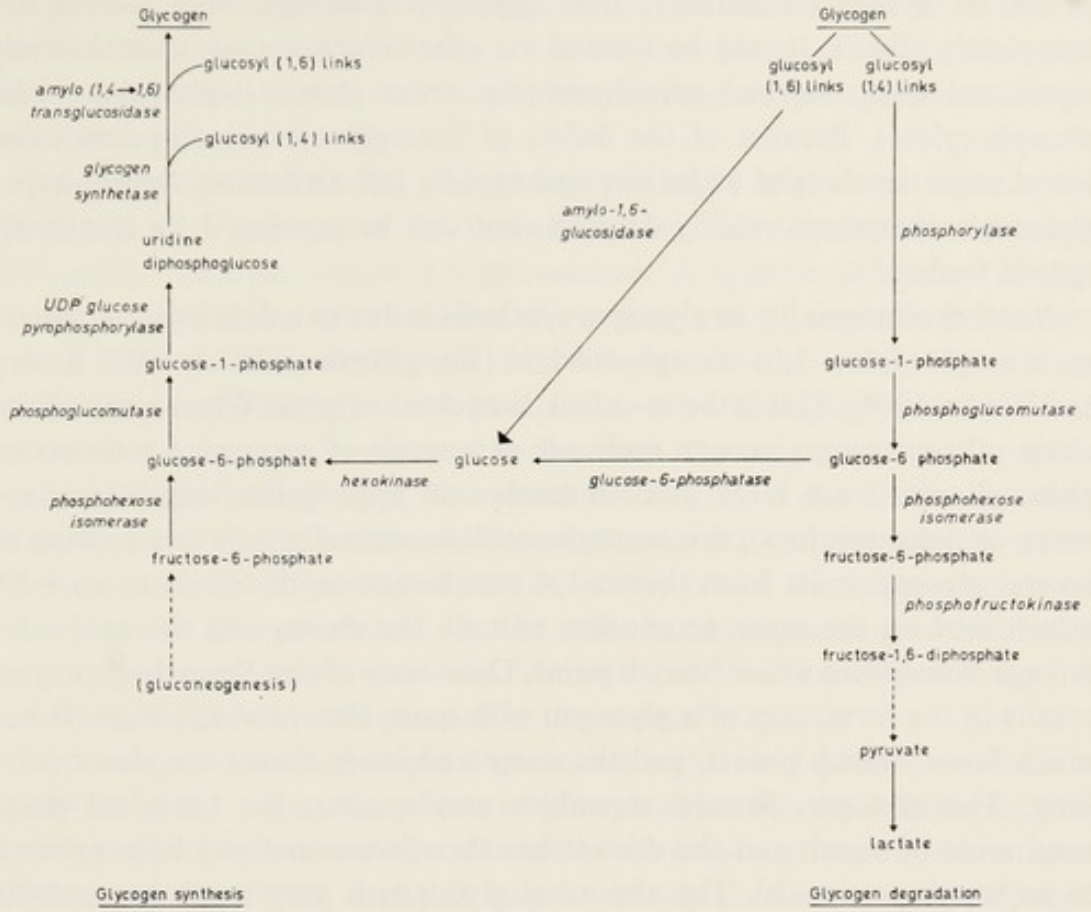


Fig. 6.9. Enzymatic steps in the synthesis and breakdown of glycogen.

6.6.1. Defects in glycogen synthesis

One abnormality of glycogen synthesis (Lewis et al. 1962, 1963) is due to a deficiency of the enzyme glycogen synthetase (UDP glucose: glycogen α-4-glucosyl transferase) which is normally concerned with the successive addition of glucose units in α-1,4 linkage to the growing ends of the peripheral glycogen chains. The glucose units are transferred from uridine diphosphoglucose (UDP glucose) which is formed from glucose-1-phosphate and uridine triphosphate (UTP) by the action of the enzyme UDP glucose pyrophosphorylase. The reactions may be written as follows:

(1) Glucose-1-phosphate + UTP → UDP glucose + pyrophosphate

(2) UDP glucose + $[\text{glucosyl (1,4)}]_n$ → UDP + $[\text{glucosyl (1,4)}]_{n+1}$

Deficiency of glycogen synthetase blocks reaction 2, and severely curtails glycogen synthesis. However it has been found that even when there is no

detectable synthetase activity, liver glycogen although very low is not completely absent. It may be formed via other reactions not quantitatively significant under normal conditions (e.g. from glucose-1-phosphate with phosphorylase). Because of the deficit of liver glycogen in this condition, blood sugar levels tend to be low and rapidly fall on fasting. Severe hypoglycaemic symptoms readily develop, but can be prevented by frequently spaced feeds.

Another abnormality in glycogen synthesis is due to a deficiency of the enzyme amylo (1,4 → 1,6) transglucosidase (Illingworth and Cori 1952, Brown and Brown 1966). This is the so-called 'brancher' enzyme. When a peripheral chain of a growing glycogen molecule as a result of successive transfers of glucosyl units from UDP glucose reaches an appropriate length (perhaps seven or more residues), this transglucosidase normally transfers a string of several glycosyl units from their α-1,4 attachment on the chain to an α-1,6 attachment on the same or another unit on the chain, and the new α-1,6 linkage constitutes a new branch point. Deficiency of this 'brancher' enzyme results in the formation of a glycogen with quite abnormal structure. It has much fewer branch points, and the outer and inner chains are abnormally long. The glycogen formed resembles amylopectin, the branched polysaccharide of starch and the disease has therefore sometimes been referred to as 'amylopectinosis'. The abnormal glycogen is very much less soluble than normal glycogen and tends to be precipitated in the tissues. Cirrhosis of the liver (Andersen 1956) is a characteristic feature of the disorder, and this has been attributed to a tissue reaction due to the abnormal glycogen being treated as if it were a foreign body.

6.6.2. Defects in glycogen mobilisation

Various enzymes are concerned in the breakdown of glycogen. The situation is however complicated by the fact that an enzyme concerned with a particular step may not be the same in all tissues. Thus the successive cleavage of the α-1,4 links in the outer chains of the macromolecule is brought about by 'phosphorylase' and leads to the formation of glucose-1-phosphate. But the phosphorylase enzyme protein in muscle is different from that in liver, and a mutation may affect one and not the other. In one form of phosphorylase deficiency, known as McArdle's disease, only muscle phosphorylase is affected and the glycogen accumulation is confined to this tissue (McArdle 1951, Schmid et al. 1959). Abnormalities confined only to liver phosphorylase have also been found (Hers 1959). These phosphorylase deficiencies are probably an even more heterogeneous group because in at least one instance

an apparent deficiency of liver phosphorylase has been shown to be due in fact to a deficiency of the enzyme phosphorylase kinase which is necessary for the conversion of the inactive dephospho form of phosphorylase to its active state (Hug et al. 1966).

Phosphorylase cleaves the α-1,4 glucosyl linkages of glycogen but cannot attack the α-1,6 linkages at the branch points. This requires the so-called 'debrancher' enzyme, amylo 1,6 glucosidase. A specific deficiency of this enzyme results in the accumulation of an abnormal glycogen with another type of unusual structure (Forbes 1953, Illingworth et al. 1956). There is an increased proportion of branch points (i.e. α-1,6 linkages), and the outer chains of the macromolecule tend to be much shorter than normal, particularly so in the fasting state. The peculiar structure is due to the fact that phosphorylase which is present in normal amounts can degrade the outer chains of the macromolecules but its action ceases as the outer tiers of branch points are approached. On the other hand lengthening of the chains by glycogen synthetase action and formation of new branch points by the brancher enzyme can still proceed in the normal way.

The main product of glycogen degradation is glucose-1-phosphate which is produced by the sequential action of phosphorylase on α-1,4 linkages. The glucose-1-phosphate is readily converted by the enzyme phosphoglucomutase to glucose-6-phosphate, and this may be metabolised via several different pathways. In the liver it is in large part hydrolysed by glucose-6-phosphatase to give rise to free glucose, which passes into the circulation, and is the main metabolic source of blood glucose. Glucose-6-phosphatase besides being present in liver also occurs in kidney, but is not found in muscle. A deficiency of this enzyme (Cori and Cori 1952) causes a characteristic form of glycogen storage disease known as Von Gierke's disease (Von Gierke 1929). In this condition glycogen accumulates in both liver and kidney which become very enlarged, but there is no accumulation in muscle. Marked hypoglycaemia is a characteristic feature, and growth tends to be severely retarded.

In muscle, glucose-6-phosphate is largely metabolised via the glycolytic pathway, most of the energy for muscular activity being derived from the breakdown of glycogen to lactate by this sequence of reactions. Phosphofructokinase which catalyses the phosphorylation of fructose-6-phosphate to give fructose-1,6-diphosphate is one of the enzymes involved, and a form of glycogen storage disease involving the gross deficiency of phosphofructokinase has been described (Tarui et al. 1965, Layzer et al. 1967). Here there is an accumulation of fructose-6-phosphate and glucose-6-phosphate as well

as of glycogen itself. Clinically, a marked weakness and stiffness occurring in the muscles on vigorous or prolonged exertion is a notable feature.

There is another form of glycogen storage disease which is of special interest because it was found to be due to the deficiency of an enzyme not originally thought to be involved in glycogen degradation at all, and because it may prove to be the prototype of a number of other inherited abnormalities in which the intracellular accumulation of some complex macromolecule is the chief feature (see p. 170). The condition is called Pompe's disease. Glycogen accumulation occurs in most tissues in the body, but it is particularly marked in the heart which becomes greatly enlarged (di Sant'Agnese et al. 1950). The patients seem to be normal at birth but the abnormality is rapidly progressive and they frequently die before the age of one year. Apart from the gross accumulation of glycogen, carbohydrate metabolism as such does not seem to be impaired, and all the enzymes known to be involved in the main pathways of glycogen degradation occur in normal amounts.

This was a very puzzling problem for many years, but eventually it was shown (Hers 1962) that the disease is in fact due to the deficiency of an α-1,4 glucosidase which normally occurs as one of a number of different hydrolytic enzymes in the cytoplasmic organelles known as lysosomes. This glucosidase, like other lysosomal enzymes, has a characteristically low *p*H optimum (about *p*H 4.0). It can hydrolyse maltose, linear oligosaccharides, and the outer chains of glycogen to give glucose. In Pompe's disease, electronmicroscopy of the liver has shown that most of the glycogen present is segregated in large vacuoles not seen in other forms of glycogenosis (Baudhuin et al. 1964), although some glycogen is also seen to be freely dispersed in the cytoplasm as occurs normally. The vacuoles appear to be lysosomes which are grossly enlarged and distended by glycogen accumulation. Other lysosomal enzymes are present in Pompe's disease, so that the lysosomal α-1,4 glucosidase deficiency appears to be specific. It seems probable that in the normal cell fragments of glycogen are constantly being taken up by the lysosomes and degraded. In Pompe's disease the uptake of glycogen fragments goes on but their degradation is blocked by the absence of the glucosidase, so that the organelles become grossly swollen with glycogen. This in turn leads to progressive degenerative changes in the cells. It is not clear why the effect although general is particularly marked in heart muscle.

6.7. Other 'storage' diseases

Besides the so-called 'glycogen storage' diseases a variety of other rare

inherited disorders have been recognised in which the outstanding feature is the progressive accumulation of particular types of complex substances in cells of various tissues. Such substances appear in many cases to represent intermediates in the normal pathways of synthesis or breakdown of compounds which form part of cellular structure and which presumably have a special role in normal function. The primary defect in most of these so-called 'storage' diseases appears to be a block at some point in the normal stepwise degradation of the compound due to the deficiency of a specific hydrolytic enzyme. Accumulation occurs because while synthesis can still go on, degradation cannot proceed further than the step normally catalysed by the deficient enzyme.

Some of the conditions in which the 'storage' compound has been characterised at least partially and in which evidence for a specific enzyme deficiency has been obtained by *in vitro* studies on tissues from affected patients, are listed in table 6.4.

In certain of these disorders (1-4, table 6.4) the substances which accumulate are lipids or glycolipids, having in common as part of their structures the moiety known as ceramide (fig. 6.10). This consists of the long chain amino-alcohol sphingosine, bound through an amide bond to a long chain fatty

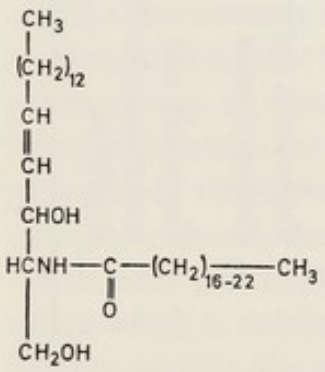


Fig. 6.10. Ceramide (N-acylsphingosine).

acid. In the various 'storage' compounds in these disorders different substituents are linked to carbon-1 of the sphingosine portion of ceramide; for example glucose in the glucoceramide which accumulates in Gaucher's disease, and phosphorylcholine in the sphingomyelin which accumulates in Niemann–Pick disease.

The characteristic features of the different diseases depend on the normal distributions of the various compounds which accumulate in an undegraded or partially degraded form, and also on the degree of the enzyme deficiency and on its normal localisation. For example at least two distinct forms of

TABLE 6.4

'Storage' diseases due to specific enzyme deficiencies.

Eponym	Enzyme deficiency	'Storage' compound	Main sites	References
1. Gaucher's disease	Glucocerebrosidase	Glucocerebroside (ceramide-glucose)	Spleen, liver, bone marrow and brain	Brady et al. (1965, 1966a), Patrick (1965), Öckerman and Köhlin (1968)
2. Metachromatic leucodystrophy	Arylsulphatase-A	Cerebroside sulphate (ceramide-galactose-3-sulphate)	Brain, kidney	Austin et al. (1963), Mehl and Jatzkewitz (1965), JatzkewitzandMehl(1969)
3. Fabry's disease	Ceramide trihexosidase	Ceramide trihexoside (ceramide-glucose-galactose-galactose)	Skin, blood vessels, brain, intestine	Sweeley and Klionsky (1963), Brady et al. (1967a, b).
4. Niemann–Pick disease	Sphingomyelinase	Sphingomyelin (ceramide-phosphorylcholine)	Spleen, liver, brain	Brady et al. (1966b), Sloan et al. (1969)
5. Generalised gangliosidosis	'Acid' β-galacto-sidase	Monosialoganglioside (fig. 6.11); galactose-rich acid mucopolysaccharides	Brain, liver	Landing et al. (1964), Van Hoof and Hers (1968), Okada and O'Brien (1968), Dacremont and Kint (1968)
6. Wolman's disease	'Acid' lipase	Triglycerides and cholesteryl esters	Liver, adrenal, spleen, lymph nodes, intestine	Patrick and Lake (1969)
7. Fucosidosis	α-fucosidase	Fucose-rich acid mucopolysaccharides and glycolipids	Liver, brain, kidney	Van Hoof and Hers (1968)

See also 'glycogen' diseases table 6.3.

Gaucher's disease probably due to different mutant genes have been recognised. These are usually referred to as the 'adult' and 'infantile' forms of the condition. In the adult form the major clinical features are enlargement of the spleen and liver and the occurrence of pathological fractures of the long bones. The central nervous system is relatively uninvolved. In the 'infantile' form there is marked damage to the central nervous system in addition to the typical features seen in the 'adult' form of the disease. Patients with the 'infantile' form suffer severe and progressive neurological damage and generally live only a year or two, whereas patients with the 'adult' form live much longer and have a quite variable life span. The substance which accumulates in the reticuloendothelial cells of the spleen, liver and bone marrow in both the 'adult' and 'infantile' forms, and in the neuronal cells of the central nervous system in the 'infantile' form is glucocerebroside which is ceramide to which is attached one glucose residue. An enzyme which hydrolyses this substance to give ceramide and glucose has been assayed in spleen samples from a series of patients with Gaucher's disease and in appropriate controls (Brady et al. 1965, 1966b; Patrick 1965; Brady 1968). It was found that in the 'adult' form of Gaucher's disease the level of the enzyme activity was on average about 15% of that found in the control subjects. However in patients with the 'infantile' form of Gaucher's disease the specific enzyme activity was virtually absent.

It seems likely that the parent substance from which the glucocerebroside accumulating in reticuloendothelial cells derives, is a complex glycolipid (globoside) which forms part of the structure of the stroma of the red blood cells (Brady 1968). As red cells age and are destroyed this material is taken up by reticuloendothelial cells where it is broken down. One of the enzymes which takes part in this stepwise degradation is evidently the glucocerebrosidase which is deficient in both forms of Gaucher's disease but to different degrees, and the rate of accumulation of glucocerebroside in spleen, liver and bone marrow is probably largely a reflection of the amount of residual activity of the deficient enzyme. However the glucocerebroside which accumulates in neurones of the central nervous system is probably derived from another substance which is a normal constituent of neuronal cells. This material, often referred to as monosialoganglioside has the structure shown in fig. 6.11. Again one of the enzymes thought to be concerned in its stepwise degradation is glucocerebrosidase. However in neuronal cells significant and damaging accumulation of glucocerebroside evidently only occurs if the enzyme is almost completely deficient as in 'infantile' Gaucher's disease, since neurological damage is not apparently a significant feature of the

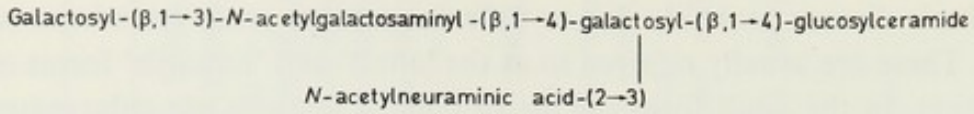


Fig. 6.11. Monosialosyl-N-tetraglycosyl-ceramide (monosialoganglioside).

'adult' form of the disease where there is only a partial deficiency of the enzyme.

Monosialoganglioside (fig. 6.11) itself is the 'storage' compound in neuronal cells in another condition: generalised gangliosidosis. Here it has been shown that there is a virtually complete deficiency of a particular type of β-galactosidase characterised by a low *p*H optimum (Van Hoof and Hers 1968, Okada and O'Brien 1968). The enzyme is presumably responsible in the normal organism for the removal of the terminal galactose unit from the ganglioside as the first step in its degradation. It may also be concerned in the normal degradation of certain complex acid mucopolysaccharides, since such substances have been found to accumulate in abnormal amounts in the liver in the same condition (Van Hoof and Hers 1968).

Hers (1965) following his discovery that the accumulation of glycogen in Pompe's disease (p. 166) is due to a specific deficiency of a lysosomal enzyme α-1,4 glucosidase, suggested that certain other 'storage' diseases might be found to have an analogous pathology. That is to say they might be due to the specific deficiency of one or another of the enzymes localised in lysosomes. Lysosomes (de Duve 1963) are cytoplasmic organelles which contain a large number of different hydrolytic enzymes most of which have relatively low *p*H optima (usually between *p*H 3.0 and *p*H 6.0). The enzymes appear to be separated from the rest of the cytoplasmic contents of the cell by a thin membrane which effectively prevents their leakage into the surrounding cytoplasm, and the organelles are thought to represent a kind of intracellular digestive system which is capable of taking up and degrading a variety of different materials which may be polysaccharide, lipid, protein or nucleic acid in nature.

Hers (1965) pointed out that in a 'storage' disease due to a specific deficiency of one of the lysosomal enzymes, one would expect to find that the substance or substances which accumulate intracellularly would be segregated in distended vesicles rather than diffused through the cytoplasm. This type of appearance has indeed been observed by electronmicroscopy of appropriate tissues in several of the conditions listed in table 6.4. Also the deficient enzyme in certain of the conditions (e.g. α-fucosidase and acid β-galactosidase) have been shown at least in rat liver to be localised to lysosomes. Thus it

seems not unlikely that at least some and perhaps most of these conditions are examples of so-called 'lysosomal' diseases.

6.8. Heterozygotes

Most of the inborn errors of metabolism are inherited as 'recessive' abnormalities. That is to say, affected individuals with the typical clinical and metabolic features of the disease usually appear to be homozygous for the particular abnormal gene, while the corresponding heterozygotes with one dose of the abnormal gene and one of its normal allele are generally quite healthy. There is a characteristic familial distribution. The disorder occurs on average in one in four of the sibs of affected individuals. It is however only rarely seen in their parents, children and other relatives. If the abnormality is sufficiently uncommon a significant increase in the incidence of cousin marriage in the parents of affected individuals may be observed.

Although apparently quite healthy the heterozygotes usually have a partial deficiency of the enzyme and often exhibit a minor metabolic disturbance qualitatively similar to that found in the affected patients. Such disturbances are usually relatively slight. They can best be studied by comparing the healthy parents and children of affected individuals with a randomly selected series of appropriate controls. These parents and children will, with rare exceptions due to fresh mutations, be heterozygotes. The randomly selected controls will if the abnormality is uncommon, almost entirely consist of homozygotes for the normal allele.

Extensive studies on these lines have been carried out on heterozygotes for the gene which in homozygotes causes phenylketonuria. Blood phenylalanine levels in patients with phenylketonuria are considerably elevated and are usually about thirty or more times greater than those found in control subjects. In the parents of phenylketonurics there is also a significant elevation of blood phenylalanine (Hsia et al. 1957, Knox and Messinger 1958), but it is very slight, and on average the blood phenylalanine levels in the fasting state are only about one and a half times those found in randomly selected controls. Furthermore there is considerable variation from individual to individual and although the difference between the average values is quite significant the two distributions overlap considerably. So discrimination of heterozygotes is incomplete and 25% or more have blood phenylalanine levels which fall within the range of values found among normal homozygotes. The minor metabolic disturbance in these heterozygotes may also be demonstrated by phenylalanine loading tests (Hsia et al. 1956). If a standard

amount of phenylalanine is given by mouth, the blood phenylalanine rises to higher levels in the heterozygotes than in the controls, and it returns to the fasting level more slowly (fig. 6.12). Also the increase in blood tyrosine which

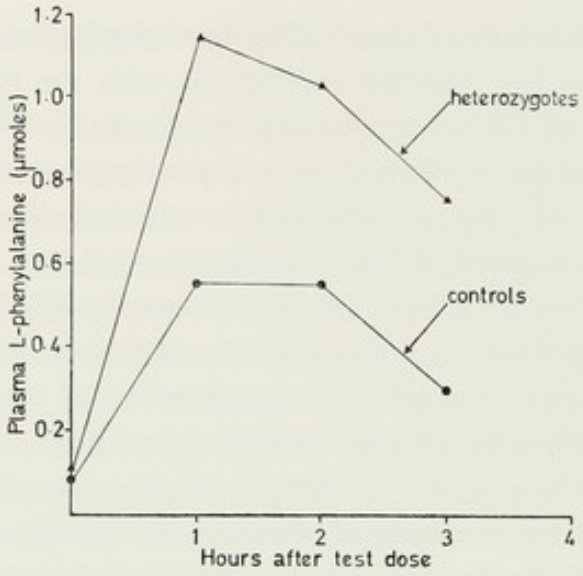


Fig. 6.12. Phenylalanine tolerance tests in heterozygous parents of phenylketonuric patients, and in appropriate controls (Hsia et al. 1956). Plasma phenylalanine levels were measured in the fasting state and at 1 hour, 2 hours and 4 hours after an oral dose of 0.1 g L-phenylalanine per kg body weight. The points plotted in the diagram are the mean values for each group of subjects at the different times. These values and the standard deviations are given below.

	Plasma phenylalanine (μmoles/ml)					
	Controls			Heterozygotes		
Time	No. tested	Mean	S.D.	No. tested	Mean	S.D.
0 (fasting)	34	0.067	0.032	37	0.103	0.029
1 hr	19	0.55	0.186	19	1.14	0.187
2 hr	19	0.55	0.168	19	1.03	0.187
4 hr	19	0.30	0.076	19	0.76	0.292

follows phenylalanine administration tends to be less marked in the heterozygotes (Jervis 1960). Similarly if phenylalanine is administered intravenously, the rate of the subsequent decline in blood level is significantly slower in the heterozygotes (Bremer and Neumann 1966). A somewhat metabolically more remote effect has also been noted after phenylalanine loads. The hetero-

zygotes have been found to excrete on average more *o*-hydroxyphenyl-lactic acid than control subjects (Berry et al. 1957).

All these results suggest that there is a partial deficiency of the enzyme phenylalanine hydroxylase in the liver in heterozygotes for the phenylketonuric gene. But although this leads to a minor disturbance of phenylalanine metabolism, it does not lead to any untoward clinical or developmental abnormality. A direct determination of the degree of enzyme deficiency has not been obtained, because phenylalanine hydroxylase occurs only in liver, and liver biopsy in otherwise healthy people is hardly justified.

Other enzymes however have a wider tissue distribution and direct enzyme assays are more feasible. For example galactose-1-phosphate uridyl transferase, the enzyme which is deficient in galactosaemia occurs in many tissues, though most of the metabolism of ingested galactose takes place in liver. Assays of the enzyme can be conveniently carried out in red cells (Kirkman and Bynum 1959, Donnell et al. 1960). In galactosaemic patients themselves virtually no activity is detected. Their healthy parents show on average in their red cells about half the level of activity of the enzyme found in randomly selected individuals. This partial enzyme deficiency has also been demonstrated in cells grown in tissue culture from small skin biopsies (Russell and DeMars 1967), so it is presumably a general phenomenon. That it occurs in liver is indicated by the results of galactose tolerance tests (Holzel and Komrower 1955). After a standard load of galactose administered by mouth, the blood galactose rises on average to higher levels in the parents of galactosaemics than in control subjects, and its subsequent rate of decline is slower.

Enzyme assay in accessible tissues such as red cells, leucocytes and skin fibroblasts grown in tissue culture, has proved to be very useful in the study of the heterozygous state for genes which determine many of the inborn errors. Quite a large number of different conditions have now been examined (table 6.5). In every case the heterozygote appears to have a partial deficiency of the enzyme which is grossly reduced or is undetectable in the abnormal homozygote. As a general rule the average level of enzyme activity in the heterozygotes is intermediate between the very low levels seen in the abnormal homozygote and the levels found in randomly selected controls. Where for example the enzyme is completely or almost completely absent in the affected homozygotes, the values found in the heterozygotes are usually about 50% of those found in normal homozygotes. Thus there often seems to be a simple gene dosage relationship. Two doses of the normal allele in the homozygote lead to the formation of twice as much enzyme as one dose of

TABLE 6.5

'Inborn errors' in which partial enzyme deficiencies have been demonstrated by *in vitro* studies on tissues from clinically unaffected heterozygotes.

Condition	Enzyme	Tissue	References
1. Histidinaemia	Histidase (histidine deaminase)	Skin	La Du et al. (1962), Holton (1965)
2. Homocystinuria	Cystathionine synthetase	Liver	Finkelstein et al. (1966)
3. Maple syrup urine disease	Branched chain ketoacid decarboxylase(s)	Leucocytes	Goedde et al. (1966b), Dancis et al. (1965), Goedde and Keller (1967)
4. Argininosuccinic aciduria	Argininosuccinase	Erythrocytes	Tomlinson and Westall (1964)
5. Hexokinase deficiency	Hexokinase (erythrocyte isozyme)	Erythrocytes	Valentine et al. (1967)
6. Phosphohexose isomerase (glucose phosphate isomerase deficiency)	Phosphohexose isomerase (glucose phosphate isomerase)	Erythrocytes	Baughan et al. (1968), Paglia et al. (1969)
7. Triosephosphate isomerase deficiency	Triosephosphate isomerase	Erythrocytes and leucocytes	Schneider et al. (1965), Valentine et al. (1966)
8. Pyruvate kinase deficiency	Pyruvate kinase	Erythrocytes	Tanaka et al. (1962), Tanaka and Valentine (1968)
9. Diphosphoglycerate mutase deficiency	Diphosphoglycerate mutase	Erythrocytes	Schröter (1965)
10. Glucose-6-phosphate dehydrogenase deficiency (primaquine sensitivity, favism etc.)	Glucose-6-phosphate dehydrogenase	Erythrocytes	See pp. 121-134
11. Galactokinase deficiency	Galactokinase	Erythrocytes	Gitzelmann (1967)

12.	Galactosaemia	Galactose-1-phosphate uridyl transferase	Erythrocytes, leucocytes and fibroblasts grown in tissue culture	Kirkman and Bynum (1959), Donnell et al. (1960), Hugh-Jones et al. (1960), Russell and De Mars (1967)
13.	Forbes' disease (glycogen storage disease type III)	Amylo-1, 6-glucosidase	Leucocytes, erythrocytes	Williams et al. (1963), Van Hoof (1967) (but see Huijing et al. 1968)
14.	Andersen's disease (glycogen storage disease type IV)	Amylo (1, 4 → 1, 6) transglucosidase	Leucocytes	Legum and Nitowski (1969)
15.	Acatalasia	Catalase	Erythrocytes	Nishimura et al. (1959), Aebi et al. (1964), Aebi (1967)
16.	Congenital methaemoglobinaemia	Methaemoglobin reductase	Erythrocytes	Scott (1960)
17.	Orotic aciduria	Orotidine-5'-phosphate decarboxylase *and* orotidine-5'-phosphate pyrophosphorylase	Leucocytes, fibroblasts grown in tissue culture	Smith et al. (1961), Fallon et al. (1964), Krooth (1964)
18.	Suxamethonium sensitivity	Serum cholinesterase (pseudocholinesterase)	Serum	See pp. 109-119
19.	Hypophosphatasia	Serum alkaline phosphatase	Serum	Currarino et al. (1957), Rathbun et al. (1961)
20.	Erythrocyte glutathione peroxidase deficiency	Glutathione peroxidase	Erythrocytes	Necheles et al. (1969)

the allele in heterozygotes, and there is usually no obvious compensation in activity of the normal allele for the defective activity of the normal allele in the heterozygote. It is, however, possible that there are exceptions to this general rule, because in some instances the limited data so far available suggests that the average level of activity seen in heterozygotes may be greater (e.g. in the Swiss type of acatalasia, Aebi et al. 1968) or smaller (e.g. in orotic aciduria, Fallon et al. 1964) than would be expected from a simple dosage relationship. If confirmed, detailed analysis of these apparently unusual situations will be of obvious interest.

Although the average level of enzyme activity in heterozygotes is generally significantly less than that in normal homozygotes, it should be noted that there is always considerable variation about the means. The two distributions usually overlap so that it is not always possible to identify the heterozygote unequivocally by determinations of enzyme level. This variation is often largely due to extraneous non-genetic factors, and the discrimination between heterozygotes and normal homozygotes can often be improved by their identification and elimination from the test system. However, it is probable that in some cases the variation is at least in part genetic in origin. It may, for example, arise because there are in fact several different so-called 'normal' alleles, each resulting in a distinct average activity level within the normal range, or different 'abnormal' alleles giving different degrees of deficiency. Another possibility is that variation in genes at other loci may affect the overall level of the particular enzyme.

The fact that heterozygotes for genes determining inborn errors of metabolism are usually perfectly healthy, implies that in the corresponding normal homozygotes the amount of enzyme present is well in excess of that required for ordinary metabolic function. Reduction to as much as half of its normal level may have no obvious pathological effects, so there is clearly a considerable functional reserve. While this is no doubt true for many and perhaps most enzymes, it may not be so for all. If the level of an enzyme without such a degree of functional reserve were reduced, perhaps to half of its normal value, then this might lead to obvious pathological consequences and clinical disorder. Heterozygotes for a gene causing a gross deficiency of such an enzyme are therefore likely to exhibit some characteristic clinical disorder. Such a disease would be expected to show a typical familial distribution. It would appear to be transmitted directly from parent to child in successive generations, and on average about half the children of matings between an affected patient and a normal individual would be affected. In other words it would appear to be inherited as a so-called 'dominant'

abnormality. The homozygotes for the abnormal gene may have a much more severe form of the disorder than the heterozygotes. However they would constitute only a small proportion of all cases, and indeed if the gene were rare, would quite possibly never be observed at all. There are of course a large number of rare genetically determined diseases which exhibit this characteristic pattern of 'dominant' heredity. Some of them may well turn out to be due to partial deficiencies of enzymes which in the normal organism are present in adequate but limited amounts, so that a reduction of up to half the normal quantity leads to overt pathological changes. Such enzymes would be expected to have little functional reserve in the normal organism and would usually be rate limiting in the particular biochemical process they subserve.

6.9. *Defects in active transport systems*

The passage of small molecular weight substances in and out of cells across cell membranes appears to depend to a considerable extent on a variety of more or less specific active transport systems. For example absorption of foodstuffs by the organism requires the transport of the various substances which are the products of gastric and intestinal digestion across the mucosal cells of the jejunum and small intestine. The processes involved are often highly specific either for only a single substance or for a small number of related substances, and they require an energy supply which is generated metabolically in the cell. Similar specific active transport systems are also known to occur in the cells lining the renal tubules. In the kidney, blood plasma is constantly being filtered by the glomeruli, so that the glomerular filtrate contains essentially all the small molecular weight constituents of blood plasma. As the fluid passes down the renal tubules, selective reabsorption of many of these substances across cells lining the renal tubule takes place. In this way many of the small molecular weight constituents of plasma are retained by the organism while others are excreted in the urine. There are also a variety of other active transport systems in the organism which are concerned with maintaining concentration gradients of particular substances across the membranes of different cells, so producing a characteristic and more or less constant intracellular distribution of metabolites.

A number of genetically determined and quite specific defects of such active transport systems have been identified. Their precise molecular basis has not yet been clarified but they seem to involve defects of particular enzymes or 'carrier' proteins, which are essentially analogous to the enzyme

deficiencies which result in blocks in intermediary metabolism in the more classical forms of inborn errors of metabolism. The best studied abnormalities involve defects in the active transport of aminoacids, but specific transport defects of other substances such as glucose (as in renal glycosuria) are also known to occur and no doubt many more remain to be identified.

6.9.1. Cystinuria

This condition was first recognised in the last century as a disorder characterised by the tendency to form renal stones composed almost entirely of the aminoacid cystine. Individuals so affected were found to excrete cystine continuously in large amounts in the urine, and the abnormality was originally included by Garrod among the conditions which he called 'inborn errors of metabolism'. At that time and indeed for many years afterwards it was thought that the disorder was a block at some point in the normal catabolism of cystine leading to accumulation of the amino-acid, and hence to its excretion in abnormal quantities in the urine. Eventually however it was shown that this is not the case, and that the abnormal urinary excretion was essentially due to a defect in the renal tubular reabsorption of cystine from the glomerular filtrate (Dent and Rose 1951). Blood plasma levels of cystine were found not to be elevated as would be expected if accumulation were occurring, but were lower than normal. Also it emerged that cystine

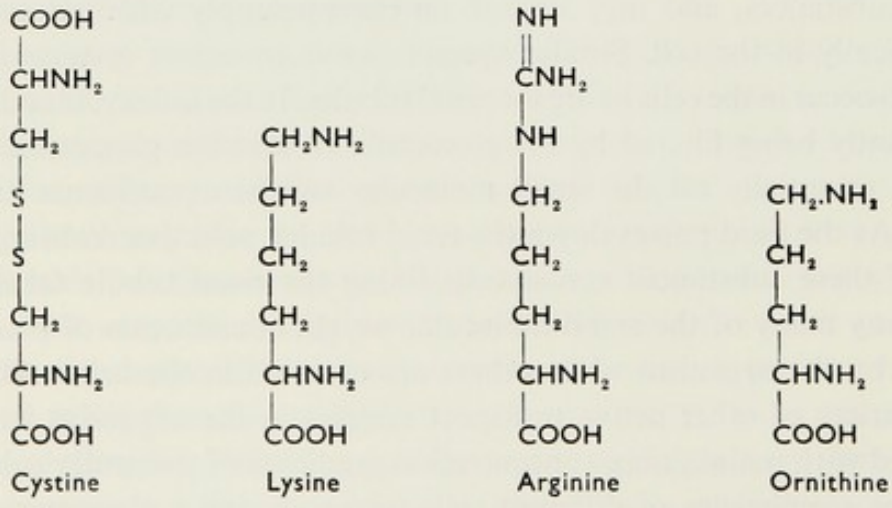


Fig. 6.13. Aminoacids characteristically excreted in excess in cystinuria.

was not the only aminoacid involved. The dibasic aminoacids lysine, arginine and ornithine (fig. 6.13), were also found to be continuously excreted in large amounts by cystinuric patients, and the renal clearances of these aminoacids

like that of cystine were greatly increased (Dent et al. 1954, Arrow and Westall 1958, Doolan et al. 1957). Thus there appeared to be a specific defect in renal tubular transport of four aminoacids normally present in blood plasma, cystine, lysine, arginine and ornithine, but not of others.

Subsequently a similar abnormality in active transport was also shown to occur in the mucosal cells of the small intestine in cystinuric patients. It results in a reduction in the rate of absorption of these aminoacids from the gut. This was first demonstrated by feeding experiments in the whole organism (Milne et al. 1961), but the specific character of the abnormality was subsequently investigated by *in vitro* studies on the uptake of these and other amino-acids by small pieces of jejunal mucosa obtained by biopsy from cystinuric and control subjects (McCarthy et al. 1964, Thier et al. 1964, 1965). These experiments confirmed the existence of a common transport defect for the four aminoacids in cystinuria. They also demonstrated competitive inhibition of the uptake of one aminoacid in this group of aminoacids by another.

It is of interest that neither the defective rates of intestinal absorption, nor the abnormal excretion in large quantities of these four aminoacids in the urine, usually lead to any nutritional abnormality in growth or development. Evidently the usual protein intake is well in excess of minimal requirements. All the clinical features of the disorder are in fact due to cystine calculus formation and the consequent obstruction of the renal tract with secondary kidney damage. The daily urinary excretion of the four aminoacids in cystinuric subjects on average amounts to about 2.0 g lysine, 1.0 g arginine, 0.75 g cystine and 0.4 g ornithine. Cystine happens to be relatively insoluble, and in urine between *p*H 5.0 and *p*H 7.0 it is kept in solution only to the extent of about 0.3–0.4 g per litre (Dent and Senior 1955). In patients who may be excreting between 0.5 and 1.0 g per day, the urinary concentration will frequently reach saturation levels, particularly at night when the urine passed is most concentrated. The cystine therefore tends to come out of solution, and this leads to calculus formation. Lysine, arginine and ornithine are, however, all very soluble and therefore do not form calculi. Because of variations in fluid intake and other physiological differences, the propensity of different cystinuric patients to form calculi, even when they are excreting similar quantities of cystine, appears to be very variable. Some may go for years without trouble, others develop renal symptoms and damage in early life.

Detailed genetical analysis of cystinuria was originally carried out by measuring the quantities of the different aminoacids in urine samples from

patients and their relatives. It soon became evident that the condition was genetically heterogenous (Harris and Warren 1953, Harris et al. 1955a, b). In one group of families, individual members showed either a grossly abnormal excretion of cystine, lysine, arginine, ornithine, or their excretion of these aminoacids was quite normal. The familial distribution indicated that the affected individuals were most probably homozygous for an abnormal autosomal gene. The presumed heterozygotes were indistinguishable from homozygotes for the normal allele. This condition was therefore called 'recessive' cystinuria.

In other families a quite different situation was found. Three types of individuals could be identified; those with a grossly abnormal excretion of cystine, lysine, arginine and ornithine; those with a moderately increased output of cystine and lysine but little or no increase in arginine and ornithine excretion; and those with normal excretion rates. Their occurrence in the families made it apparent that individuals with greatly increased output of all four aminoacids were likely to be homozygous for an abnormal gene, while individuals of the intermediate type with moderately increased cystine and lysine excretions were heterozygotes. This condition was referred to as 'incompletely recessive' cystinuria.

Thus it seemed that at least two different abnormal genes could cause cystinuria. The two sorts of homozygote showed on average a similar degree of abnormality in urinary aminoacid excretion and were indistinguishable phenotypically. However, while heterozygotes for one of the genes showed no obvious abnormality, heterozygotes for the other displayed a partial defect in renal tubular aminoacid reabsorption. The quantities of cystine and lysine excreted in heterozygotes of the 'incompletely recessive' type are very variable though closely correlated. Occasionally the cystine levels may be sufficiently high as to lead to calculus formation. But this is unusual, whereas in both sorts of abnormal homozygotes it is a frequent occurrence.

Studies of the active uptake of these aminoacids in jejunal biopsy material obtained from different patients have revealed further complexity in the genetics (Rosenberg et al. 1966, Rosenberg 1966). Three distinct types of homozygotes have been defined by such investigations (table 6.6). Type I corresponds to the homozygote for so called recessive cystinuria, because the corresponding heterozygotes show no abnormality in urinary aminoacid excretion. Types II and III correspond to homozygotes for what was previously called 'incompletely recessive' cystinuria. In Type I virtually no uptake of the specific aminoacids can be detected in the jejunal material. In Type II there is slight activity, and in Type III there is a significant degree of uptake

TABLE 6.6

Distinction between three types of cystinuria based on the degree of active transport of cystine, lysine and arginine observed by *in vitro* studies on jejunal biopsies from homozygotes, and on the urinary excretion of cystine and lysine observed in the corresponding heterozygotes.

	Active intestinal transport in homozygotes			Excretion of cystine and lysine in the urine in heterozygotes
	Cystine	Lysine	Arginine	
Controls	7.0±1.4	11.2±1.6	28.3±1.3	Normal
Cystinurics				
Type I	1.1±0.2	1.0±0.3	0.9±0.2	Normal
Type II	2.4±0.2	1.0±0.2	Not tested	Increased
Type III	4.1±2.8	4.2±3.0	6.6±3.3	Increased

The degree of intestinal active transport is expressed as a distribution ratio. A value of 1.0 indicates the absence of active transport of the particular aminoacid. Values greater than 1.0 indicate the occurrence of active transport of varying degrees (for details see Rosenberg et al. 1966, Rosenberg 1966).

though on average less than in non-cystinuric controls. Types II and III also differ in the degree of aminoaciduria shown by the corresponding heterozygotes. In Type II this is somewhat more marked than in Type III.

Family studies support the idea that at least three distinct abnormal genes are indeed involved. Each appears to affect the active transport process in a specific way. Furthermore, in several instances it has been possible to show by family studies that certain individuals with gross excretion of all four aminoacids are probably doubly heterozygous for different combinations of these abnormal genes (e.g. I-II, I-III and II-III). This suggests that the genes may be allelic and each alters but in a different manner the structure of a single enzyme or carrier protein. However, the nature of this postulated protein is not known, and it may well be that if it contained two non-identical polypeptide chains its functional properties could be modified by mutations at more than one locus. One point of particular interest, though at present unexplained, is that the gene for Type I cystinuria, which apparently in homozygotes produces the most marked transport deficit in the jejunum, does not in heterozygotes lead to any obvious renal abnormality; whereas the genes for Type II and Type III cystinuria, which have apparently a less

profound effect in the jejunum in homozygotes, cause significant renal abnormalities in heterozygotes. Also the renal defect, as indicated by the abnormal aminoacid excretion, does not obviously differ between the three types of homozygotes.

6.9.2. Other aminoacid transport defects

Another genetically determined transport defect has been shown to involve specifically the aminoacids glycine, proline and hydroxyproline (Scriver 1968, Rosenberg et al. 1968). All three aminoacids are excreted in abnormally large amounts in the urine of affected homozygotes evidently because of a specific abnormality in renal tubular reabsorption (table 6.7). There appear

TABLE 6.7

'Hereditary renal iminoglycinuria'. Estimated percentage tubular reabsorption of proline, hydroxyproline and glycine from glomerular filtrate in heterozygotes and homozygotes (Scriver 1967).

	Percentage renal tubular reabsorption from glomerular filtrate		
	Normal	Heterozygote	Homozygote
Proline	99.8	99.8	80
Hydroxyproline	100	100	65
Glycine	93	84	65

however to be no untoward clinical consequences. In heterozygotes an increased renal clearance and an abnormal excretion of only glycine is found. The findings suggest that in the tubular reabsorption of aminoacids from the glomerular filtrate in the kidney, these three aminoacids share a common transport system. It seems that some defect in an enzyme or 'carrier' protein specific to the pathway is likely to be the cause of the particular abnormality. Defective intestinal absorption of proline has also been observed in this condition (Goodman et al. 1967).

A further and somewhat larger group of aminoacids appear to share a common transport pathway in the cells of the renal tubules and probably also in the intestinal mucosa. They include alanine, serine, threonine, asparagine, glutamine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan,

histidine and citrulline. They are all excreted in excess in the urine in a condition known as Hartnup disease, in which there is presumably some defect in this common transport system.

Hartnup disease (Jepson 1965) was originally described under the title 'hereditary pellagra-like skin rash with temporary cerebellar ataxia, constant renal aminoaciduria and other bizarre biochemical features' (Baron et al. 1956). This is an apt description, though the symptomatology is somewhat variable from case to case. The other biochemical features referred to are the urinary excretion in varying amounts of certain tryptophan derivations, notably indoxyl sulphate and indolylacetic acid. These are probably derived from excessive bacterial breakdown of tryptophan in the gut consequent on its delayed absorption (Milne et al. 1961). It is possible also that the various clinical features of the disease may be caused by toxic products of bacterial degradation of tryptophan or other aminoacids in the gut, but their detailed causation is not understood. Individuals with Hartnup disease, although variable in symptomatology, all show the same very characteristic pattern of aminoacid excretion. Their distribution in families indicates that they are homozygous for a rare autosomal gene, but so far no detectable abnormality has been found in presumptive heterozygotes.

Thus these various abnormalities define three different aminoacid transport systems in the kidney and probably also in the gut. These are concerned with distinct groups of aminoacids which comprise cystine, lysine, arginine and ornithine; glycine, proline and hydroxyproline; and the large group of mainly mono-amino mono-carboxylic aminoacids excreted in excess in Hartnup disease. Each of these systems can evidently be specifically blocked (at least partially) by an appropriate mutation. There is however, evidence for the kidney that other systems concerned with the transport of some of these aminoacids also occur (Rosenberg et al. 1967, Scriver 1967, Scriver and Hechtman 1970). These systems may be specific for only one (or some) of the aminoacids in the three group transport systems mentioned above, and are probably affected by mutations at other loci. They probably differ in their capacities and in their kinetics, so that any particular aminoacid may, to a greater or less extent according to the particular conditions, be actively transported by more than one system. Thus a variety of effects can be expected from different mutations.

6.10. 'Inborn errors' of drug metabolism

6.10.1. Enzyme deficiencies and pharmacological aberrations

It has long been recognised that the administration to certain individuals of particular drugs in normal doses sometimes results in a markedly abnormal response, which may have very undesirable clinical consequences. Certain of these so-called drug idiosyncrasies have been shown to be due to specific inherited enzyme defects, essentially similar to those occurring in the more classical types of inborn errors of metabolism.

The enzyme involved may be concerned in the normal metabolism of the drug, and its defective action can result in the drug, or one of its pharmacologically active derivatives, persisting at much higher concentrations in the body than is normally the case, so that the effects of a standard dose are excessively prolonged. A well known example of this, which has already been discussed (pp. 109–119), is the abnormally prolonged period of respiratory paralysis which occurs in certain individuals when the drug suxamethonium is administered to obtain muscular relaxation. Serum cholinesterase is the defective enzyme, and it has been shown that the excessive sensitivity to the drug is due to particular alleles which result either in the synthesis of an abnormal form of the enzyme with unusual kinetic properties, so that it is not capable of hydrolysing the suxamethonium at an appreciable rate under the conditions that are obtained after the drug's administration, or which result in a complete deficiency of the enzyme. In these circumstances the drug persists in the unhydrolysed and pharmacologically active form at relatively high concentrations for a prolonged period.

In other cases the specific enzyme defect may be the cause of the drug idiosyncrasy because it sets up an abnormal metabolic situation in particular cells, so that they are not capable of dealing effectively with certain secondary pharmacological effects of the drug which under normal metabolic conditions are not harmful. Examples of this kind of phenomenon are provided by the series of structurally abnormal forms of glucose-6-phosphate dehydrogenase which in different ways give rise to a marked deficiency of glucose-6-phosphate dehydrogenase activity in red cells (pp. 121–134). In the commoner types of glucose-6-phosphate dehydrogenase deficiency there is generally no clinical abnormality under ordinary conditions, even though red cell metabolism must be in some degree disturbed. However, when particular drugs, such as the antimalarial compound primaquine, or certain sulfonamides (see also table 5.5, p. 123) are administered, the unusual metabolic state of the red cells becomes accentuated and premature destruction causing haemolytic

anaemia ensues. It is of interest that similar abnormal reactions to this group of drugs have also been observed in certain other genetically determined red cell metabolic defects. For example, primaquine and several of the other drugs which induce haemolysis in glucose-6-phosphate dehydrogenase deficiency have also been found to have a similar effect in individuals with a specific deficiency of the red cell enzyme glutathione reductase (Waller 1968).

6.10.2. Acute intermittent porphyria and porphyria variegata

The metabolic disorders known as acute intermittent porphyria (Waldenström 1937) and porphyria variegata (Dean 1963), provide another very striking illustration of how individuals with a specific inherited abnormality may react in a grossly abnormal manner when one of a group of commonly used drugs (in this case the barbiturates) are administered, usually for reasons unconnected with the metabolic disease from which they suffer.

Acute intermittent porphyria is sometimes referred to as the Swedish type of porphyria because, although relatively rare, a considerable number of examples of the condition have been identified in Sweden and have been extensively studied (Waldenström 1957, Waldenström and Haeger-Aronsen 1967). The condition, unlike most other inborn errors of metabolism, is inherited as a Mendelian 'dominant' characteristic, and the affected individuals may be presumed to be heterozygous for a specific abnormal gene. The clinical manifestations are extremely variable, but characteristically there are periodic attacks of a peculiarly severe intestinal colic, and neurological disturbances which result in irregularly distributed though often widespread paralysis and in some cases mental confusion. The acute attacks vary in severity and either the abdominal or neurological features may predominate. They do not usually occur until adult life, and their occurrence is frequently closely correlated with the taking of barbiturates or certain other drugs.

Porphyria variegata appears to be particularly prevalent in S. Africa among the white population who are descendants of the small group of Dutch settlers who came to the country at the end of the seventeenth century (Dean 1963). The condition, like acute intermittent porphyria, is inherited as a Mendelian 'dominant' characteristic, but the mutant allele for which the affected individuals are heterozygous is probably distinct since there are certain characteristic differences between the two conditions. Patients with porphyria variegata, for example, frequently show skin lesions due to hypersensitivity to sunlight, and to minor mechanical trauma, and this dermatological

abnormality is not seen in acute intermittent porphyria. However, the two conditions resemble one another in the irregular occurrence of acute abdominal and neurological episodes which are very similar in their manifestations, and in both conditions are particularly likely to be precipitated by barbiturates.

The characteristic biochemical finding in both these disorders is the appearance in the urine of grossly abnormal quantities of δ-aminolaevulinic acid and porphobilinogen, which are normal intermediates in the biosynthesis of haem (fig. 6.14). These substances are formed in excessive quantities in the

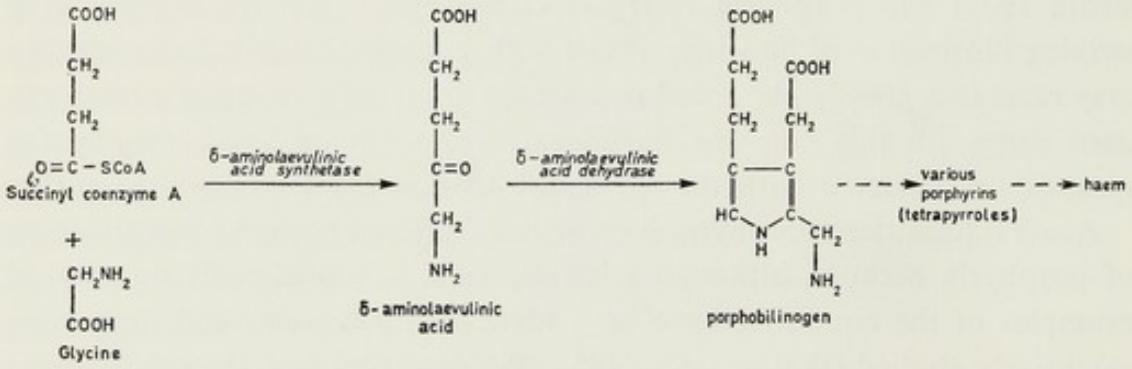


Fig. 6.14. Early stages in the biosynthesis of haem.

liver, particularly during the acute attacks. They leak into the circulation and are then excreted into the urine. In between attacks a moderately increased excretion of these metabolites is seen in acute intermittent porphyria, and they may also be found in symptom-free carriers of the gene who have not yet developed any clinical abnormality. In porphyria variegata abnormal urinary excretion of these substances is much less marked in between acute attacks and often not detectable. However, in porphyria variegata, but not in acute intermittent porphyria, the continuous excretion of abnormal amounts of protoporphyrin and coproporphyrin in the stools is a characteristic finding, and this is frequently observed in the symptom-free condition (Barnes 1958).

The precise nature of the metabolic abnormality in these disorders is still very imperfectly understood. δ-Aminolaevulinic acid and porphobilinogen are apparently formed in excessive amounts in the liver cells, but probably not in the haematopoietic cells of the bone marrow. It has been shown that the level of activity of the mitochondrial enzyme δ-aminolaevulinic acid synthetase, which normally catalyses the formation of δ-aminolaevulinic acid from glycine and succinyl-coenzyme A, is very considerably elevated in liver

samples obtained from patients with acute intermittent porphyria (Tschudy et al. 1965, Nakao et al. 1966, Dowdle et al. 1967) and a similar elevation of δ-aminolaevulinic acid synthetase has also been observed in the liver in porphyria variegata (Dowdle et al. 1967). This finding is of considerable interest if only because it is the first example in an inborn error of metabolism of a gross increase of the level of a specific enzyme, and it contrasts strikingly with the enzyme deficiencies which are usually found. The increased level of the enzyme can account for the excessive amounts of δ-aminolaevulinic acid and porphobilinogen which are synthesised in the livers of these patients, though it is far from clear how the various abdominal and neurological disorders are brought about.

The finding of elevated levels of δ-aminolaevulinic acid synthetase activity in the liver in such heterozygous conditions has led to the suggestion that the abnormal genes concerned may represent mutations at an 'operator' locus normally concerned with controlling the rate of synthesis of this enzyme (Tschudy et al. 1965, Granick 1966, Perlroth et al. 1966). However, other possibilities can be envisaged, and at present there is insufficient information to distinguish between them critically. It is of interest to note that experimental work using liver material from various animal species has shown that increased synthesis of δ-aminolaevulinic acid synthetase may be brought about by a variety of substances, some of which are known to precipitate acute episodes in intermittent porphyria and porphyria variegata in man. So eventually it may be possible to explain the pronounced effect of the particular drugs in these conditions in terms of their interaction with an already defective system concerned with the regulation of the synthesis of this enzyme. However, the nature of such possible interactions is at present very obscure.

6.10.3. Isoniazid inactivation

In some cases genetically determined differences in the activity of a certain enzyme may result in quite marked differences between individuals in the manner in which they metabolise a particular drug, although this may not be associated with any acute clinical consequences. An illustration of this is provided by the differences which have been found to occur in the acetylation of the drug isoniazid, which is widely used in the chemotherapy of tuberculosis. In its acetylated form, isoniazid is much less active therapeutically and is also less toxic, so that the drug is effectively inactivated by acetylation.

Shortly after the use of isoniazid in tuberculosis therapy was introduced, it was found that there are considerable differences between individuals in the

rate at which isoniazid is inactivated by acetylation, though in any one individual the metabolism of the drug appears to be remarkably constant (Hughes et al. 1954, Bell and Riemensnider 1957). It was then shown that people can be differentiated into two more or less sharply distinct groups according to the rate at which the inactivation proceeds. For example, if a standard dose of isoniazid is administered to a randomly selected series of individuals and the blood level measured a few hours later, the distribution of blood levels is clearly bimodal (fig. 6.15). Individuals can be readily classified into two

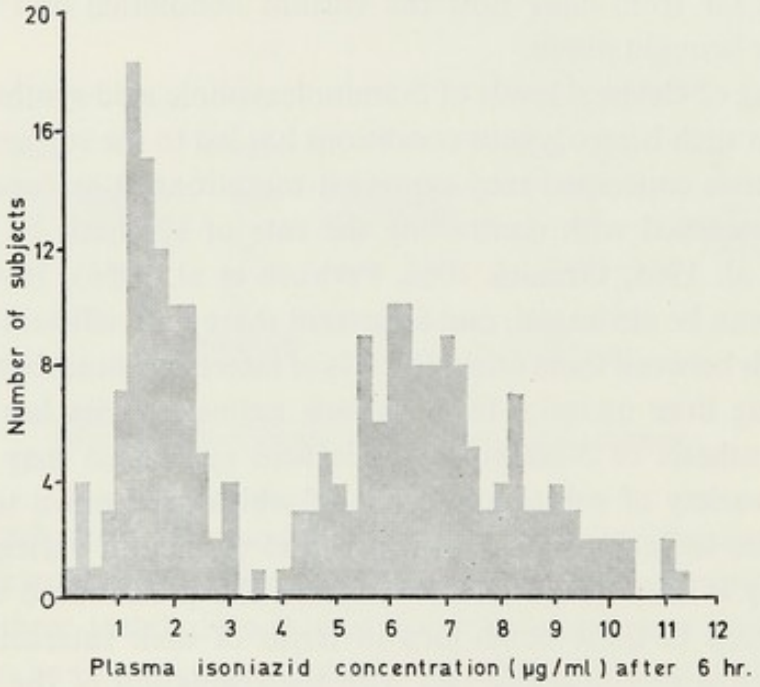


Fig. 6.15. Distribution of plasma isoniazid levels in 220 different subjects six hours after an oral dose of isoniazid (Evans et al. 1961). In this study the dose of isoniazid administered was 40 mg/kg of 'metabolically active mass' ($Wt^{0.7}$).

distinct groups: 'rapid inactivators' in whom the blood level is relatively low a few hours after taking the drug, and 'slow inactivators' in which it is relatively high (Knight et al. 1959, Evans et al. 1960, 1961). In the 'rapid inactivators' a much higher proportion of the drug is excreted in the acetylated form in the urine than in the 'slow inactivators', in whom it is mainly excreted unacetylated.

Family studies demonstrate that the differences are genetically determined, and the results can largely be accounted for in terms of two common alleles, such that 'slow inactivators' represent the homozygotes for one allele, and 'rapid inactivators' are either heterozygotes or homozygotes for the other

(Knight et al. 1959, Evans et al. 1960). It is probable that the rate of inactivation of the drug is somewhat more rapid in the homozygous 'rapid inactivators' than in the heterozygotes.

Acetylation of isoniazid is brought about by an acetyltransferase enzyme which occurs in the liver and which is concerned in a reaction by which an acetyl group from acetyl-coenzyme A is transferred to isoniazid (fig. 6.16).

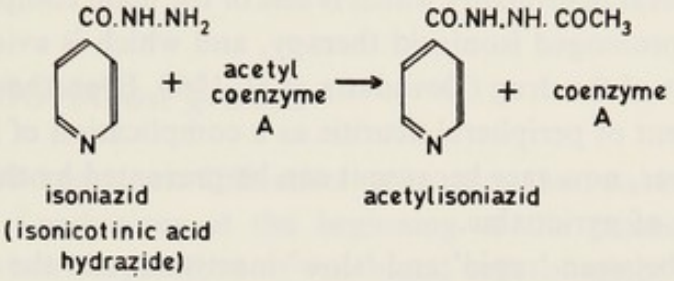


Fig. 6.16. Acetylation of isoniazid.

Assay of acetyltransferase activity in liver samples obtained by biopsy has shown striking differences in level between 'rapid' and 'slow' inactivators, the activity being on average much higher in the former group than in the latter (Evans and White 1964). Similar results have also been obtained with autopsy specimens (Jenne 1965). Semipurified preparations of the enzymes obtained from 'rapid' and 'slow' inactivators appeared to be closely similar in a number of properties, such as Michaelis constants and substrate specificities (Jenne 1965), which suggests that the difference between the two types may depend on the amount of the enzyme protein actually present in the liver cells, rather than on differences in its specific activity.

The enzyme has also been shown to be concerned in the acetylation of certain other drugs, such as sulfamethazine and hydralazine (Evans and White 1964). However, other compounds normally acetylated in the body, such as sulfanilamide (Evans and White 1964) or p-aminosalicylic acid (Jenne 1965) are not apparently dealt with by the same enzyme. These *in vitro* findings correlate with the observation that subjects receiving sulfamezathine show significant differences in the proportions of the free and acetylated forms of the drug excreted in the urine according to whether they are 'slow' or 'rapid' inactivators of isoniazid (Evans and White 1964), but the same kind of differences are not found in individuals receiving sulfanilamide (Peters et al. 1965) or p-aminosalicylic acid (Jenne 1965).

The existence of these marked individual differences in isoniazid inactivation

appears to be of little or no consequence in the therapeutic use of the drug in tuberculosis. In comparisons of large groups of patients on standardised anti-tuberculous treatment including isoniazid, significant differences have not usually been found between the results of treatment in 'rapid' and 'slow' inactivators (Evans 1963). If such differences do exist they are likely to be relatively small and of little importance in practice. But 'slow' inactivators of isoniazid appear to be somewhat more likely than 'rapid' inactivators to develop peripheral neuropathy which is one of the main complications which may occur in prolonged isoniazid therapy, and which is evidently due to a toxic side effect of the drug (Devadatta et al. 1960, Evans and Clarke 1961). The development of peripheral neuritis as a complication of isoniazid treatment is, however, now rare because it can be prevented by the simultaneous administration of pyridoxine.

Differences between 'rapid' and 'slow' inactivators in the toxic or therapeutic effects of other drugs acetylated by the same enzyme have not yet been very extensively studied. However, in one case, the use of phenelzine (β-phenylethylhydrazide) in the treatment of depression, it appeared that severe adverse effects of the drug were somewhat more common among 'slow' inactivators of isoniazid than among 'rapid' inactivators, though the therapeutic response appeared to be much the same in the two groups (Evans et al. 1965).

It is of some interest that the proportion of 'rapid' and 'slow' inactivators of isoniazid has been found to vary quite widely in different human populations. Among Europeans and Negroes about half the population appear to be 'slow' inactivators (Evans et al. 1960), whereas among Japanese (Sunahara et al. 1961) the proportion is much less – about 10%. Thus among Europeans and Negroes it appears that the allele causing a deficiency of the enzyme must be about twice as common (allele frequency ca. 0.7) as the allele responsible for relatively high enzyme activity (allele frequency ca. 0.3). This of course is in striking contrast to the situation with genes causing analogous enzyme deficiencies in the more classical types of 'inborn errors of metabolism' which are, in general, extremely infrequent.

The blood group substances

7.1. The ABO blood groups

The first inherited antigenic differences to be recognised as such in man were discovered by Landsteiner at the beginning of the present century. He showed that when suspensions of red blood cells obtained from different people are mixed with blood serum obtained from other people clear-cut differences in reaction are observed. In some cases there is marked agglutination or clumping of the red cells. In other cases the red cells remain unaffected. The agglutination is due to the binding of particular antigenic substances present on the surface of the red cells with specific antibodies (immunoglobulins) present in the serum. By cross-agglutination tests using red cells and sera from normal healthy individuals, it was found possible to classify people into four distinct groups in terms of two antigenic specificities (A and B). Some people (group O) have neither of these specificities, others have only one (group A or group B), while still others have both (group AB). The corresponding serum antibodies are called anti-A and anti-B, and their occurrence in sera of individuals of the four groups are indicated in table 7.1.

TABLE 7.1

The ABO blood groups.

Blood group	Antigenic specificities on red cells	Antibodies in serum
O	—	anti-A and anti-B
A	A	anti-B
B	B	anti-A
AB	A and B	—

These findings laid the foundation for modern blood transfusion, by defining who would be compatible donors of blood for particular recipients. Among Europeans about 47% of people are group O, about 42% group A, about 8% group B and about 3% group AB. However, the relative frequencies of the four groups varies from population to population.

Early studies on the familial distributions and the population frequencies of these four groups showed that they were inherited, and led to the hypothesis that they are determined by three allelic genes; allele *A* determining A specificity, allele *B* determining B specificity, and allele *O* being inactive. According to this, group O individuals are all homozygous *OO* and group AB individuals all heterozygous *AB*. But group A individuals may be either homozygous *AA* or heterozygous *AO*; and group B individuals may be either homozygous *BB* or heterozygous *BO*. Later work substantially confirmed this interpretation, and extended it in certain ways. For instance it was shown that by using appropriate antisera group A could be subdivided into two relatively common types A_1 and A_2, evidently determined by separate alleles, and also a number of rare alleles resulting in slightly modified specificities were identified (for detailed review see Race and Sanger 1968).

In the red cell, the substances which carry these A and B antigenic specificities are firmly bound and cannot be extracted from the stroma with water or salt solutions. Active preparations may however be obtained by extraction with ethanol. Substances with similar properties and specificities have also been shown to occur elsewhere in the body (Hartman 1941), notably in membranes of the endothelial cells which line the cardiovascular system (Szulman 1964, 1966). These substances are usually referred to as 'alcohol soluble' group specific substances. This is to distinguish them from another class of substances which carry the same antigenic specificities evidently determined by the same alleles (*A*, *B*, etc.), but which are water soluble. The so called 'water soluble' group specific substances occur in large amounts in mucous secretions, notably in saliva and the mucus of the gastrointestinal tract. They are usually demonstrated and assayed by absorption techniques. That is by showing that they have the capacity when mixed with an appropriate antiserum, to combine specifically with blood group antibody so that agglutination fails to take place upon subsequent addition of red cells of the corresponding type.

The 'alcohol soluble' group specific substances extracted from red cells appear to be glycolipids (Koscielak 1967, Hakomori and Strycharz 1968). They are complex macromolecules containing a carbohydrate moiety joined through sphingosine to fatty acids. The 'water soluble' group specific sub-

stances on the other hand are glycoproteins (Kabat 1956, Morgan 1967), containing a high percentage of carbohydrate. Thus there are two quite different classes of macromolecule which exhibit similar or identical antigenic specificities determined by the same alleles. The antigenic specificities appear to be a reflexion of the structural arrangement of certain sugars on the more superficial parts of the carbohydrate moieties of the macromolecules. These groupings are probably very similar if not identical in the two classes of macromolecules.

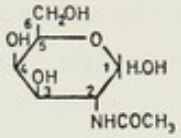
In fact most of the work on the chemical structures which underlie these antigenic specificities has been carried out on the so called 'water soluble' group specific substances. This is because until very recently these were the only form of blood group substance which could be isolated in a satisfactory state and in adequate quantities for structural studies.

The most potent sources of water soluble blood group substances among the normal secretions of the body are saliva and gastric juice. Meconium, the first stool of the newborn is also rich in these substances. However a particularly useful source of material for the isolation of relatively large amounts of these substances from single individuals has proved to be fluid obtained from ovarian cysts (Morgan and van Heyningen 1944). These fluids accumulate in the cysts over long periods of time, and large volumes often containing several grams of active group specific material may be obtained from a single cyst.

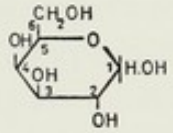
Purified blood group substances obtained from such secretions turn out to be high molecular weight glycoproteins. The molecular weights range from 3×10^5 to 1×10^6, and it seems that a preparation having specific blood group activity obtained even from a single individual and a single secretion, may contain a family of macromolecules varying somewhat in overall size and perhaps in composition, though they are no doubt very closely related in structure. Such substances usually contain about 85% carbohydrate and 15% aminoacids. The full details of their molecular organisation are not yet known, but their general properties and degradation products suggest that they are made up of a large number of relatively short oligosaccharide chains which are covalently attached at intervals to a polypeptide backbone. On the assumption that the specific substances usually have a molecular weight around 500,000, and that the carbohydrate chains are composed of seven or eight sugar units, there would appear to be some 300 carbohydrate chains, each with a non-reducing end group (Morgan and Watkins 1969).

The carbohydrate moiety of the macromolecules contains (fig. 7.1) a hexose, D-galactose, a methyl pentose, L-fucose and two aminosugars, N-acetyl-D-glucosamine and N-acetyl-D-galactosamine. The nine carbon

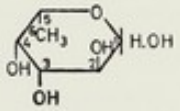
N-acetyl-D-galactosamine



D-galactose



L-fucose



N-acetyl-D-glucosamine

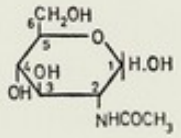


Fig. 7.1. Sugars which occur in the blood group substances.

sugar, N-acetyl-neuraminic acid (sialic acid) is also often present. The polypeptide part of the macromolecules is composed of fifteen aminoacids, and is distinctive in that four aminoacids, threonine, serine, proline and alanine make up about two thirds of the aminoacids present (Pusztai and Morgan 1963). Another notable feature is that sulphur containing aminoacids are virtually absent. The integrity of the complete macromolecule is essential for maximum serological reactivity, and the role of the polypeptide backbone appears to be that of maintaining the correct spacing and orientation of the carbohydrate chains which provide in their terminal groupings the determinants of the antigenic specificity.

The terminal sequences of sugars which are now thought to confer A and B group specificity are shown diagrammatically in fig. 7.2 (Watkins 1966). The essential point is that the A specific chains terminate with an N-acetylgalactosamine residue, while the B specific chains terminate in a galactose residue. In other respects the chains are the same, although it should be noted that both A and B specificity can occur on chains in which there may be a β-1,3 (type 1 chains) or β-1,4 (type 2 chains) link in the second position between a galactose and an N-acetylgalactosamine residue. Also shown in fig. 7.2 are oligosaccharide chains which occur in many of these group specific substances and give rise to what has been called H specificity. It will be noted

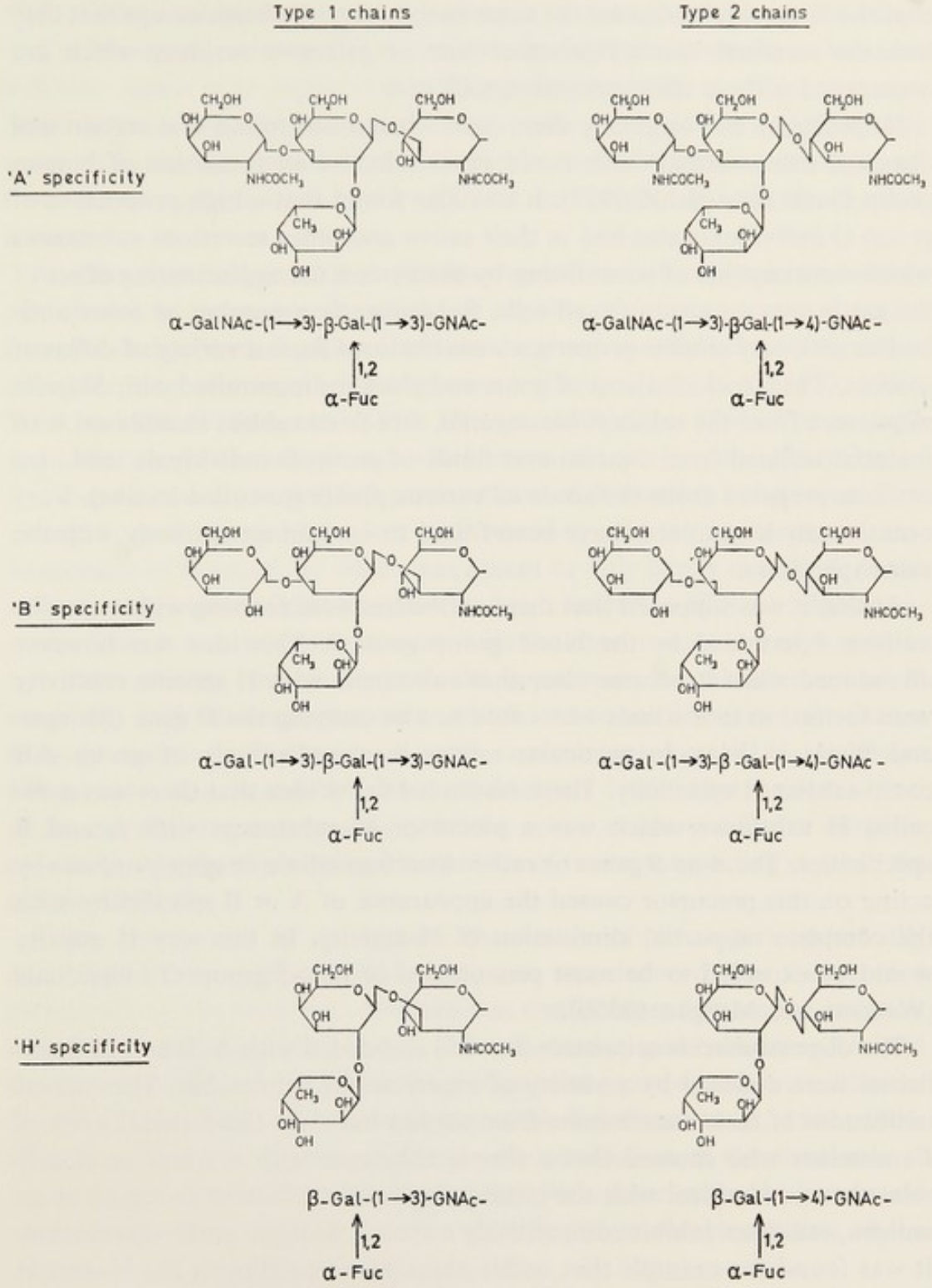


Fig. 7.2. Terminal sugar sequences in polysaccharide chains of glycoproteins which confer 'A', 'B' and 'H' specificity (see text). Gal: D-galactopyranosyl; Fuc: L-fucopyranosyl; GNAc: N-acetyl-D-glucosaminopyranosyl; GalNAc: N-acetyl-D-galactosaminopyranosyl.

that the H specific chains are the same as the A and B chains except that they lack the terminal N-acetylgalactosamine or galactose residues which are necessary for these characteristic specificities.

H specificity was originally discovered when it was found that certain sera obtained from normal cattle could cause selective agglutination of human group O red cells (Schiff 1927). It was also found that a high proportion of group O individuals also had in their saliva and other secretions substances which were capable of neutralising by absorption the agglutinating effect of the cattle sera on group O red cells. Subsequently a number of other antibodies with very similar properties were obtained from a variety of different sources. These included sera of goats and chickens immunised with *Shigella shiga*, sera from the eel *Anguilla anguilla*, sera from rabbits immunised with material isolated from ovarian cyst fluids of group O individuals, and also extracts prepared from the seeds of various plants (so-called lectins). Very occasionally human sera have been found to contain an antibody with the same specificity.

At first it was supposed that these antibodies were reacting with a specific antigen determined by the blood group gene *O*. This idea was however abandoned when it became clear that substances with H specific reactivity were formed in individuals who could not be carrying the *O* gene (Morgan and Watkins 1948). In particular salivas from individuals of group AB could exhibit H specificity. These results led to the idea that there was a so-called H substance which was a precursor of substances with A and B specificities. The *A* or *B* genes or rather their immediate enzyme products by acting on this precursor caused the appearance of A or B specificities with the complete or partial elimination of H activity. In this way H activity would be expected to be most pronounced in blood group O individuals (Watkins and Morgan 1955b).

The oligosaccharide sequences (fig. 7.2) associated with A, B and H specificities were deduced by a variety of experimental approaches. The earliest indications of their nature came from studies based on the classical work of Landsteiner who showed that a simple substance with a structure closely related to or identical with the immunological determinant grouping of an antigen, can often inhibit competitively a specific antigen–antibody reaction. It was found for example that under appropriate conditions the H–anti H reaction could be inhibited by L-fucose, but not by other sugars present in the group specific glycoproteins. Similarly the A–anti A reaction could be specifically inhibited by N-acetylgalactosamine and the B–anti B reaction by D-galactose. These findings (Watkins and Morgan 1952, Morgan and Watkins

1953, Kabat and Leskowitz 1955) indicated that despite the overall similarities in the apparent composition of the A, B and H group specific substances different sugars were implicated in the immunologically specific groupings. This approach was further extended and made more specific and also more sensitive by the use as inhibitors of particular di- and trisaccharides obtained from partial hydrolysates of the different substances, and also from other sources.

Another approach came from the discovery of certain enzymes in a number of different microorganisms, which are capable of splitting terminal sugars from the ends of carbohydrate chains in the group specific substances, and at the same time altering their antigenic specificity (Iseki et al. 1953, 1959; Watkins 1956; Watkins et al. 1962; Harrap and Watkins 1964). For example particular enzymes were found which cause the preferential liberation of free D-galactose from purified preparations of B reacting substances. This change is accompanied by a loss of B reactivity and the appearance or enhancement of H reactivity which was absent or only barely detectable in the original preparation. Other enzymes lead to a liberation of N-acetylgalactosamine from A reacting glycoproteins, with the concomitant loss of A reactivity. Here also the change leads to the appearance or enhancement of H reactivity. Still other enzymes occur which lead to the liberation of L-fucose from H reacting substances, and the concomitant loss of H reactivity. Furthermore it is found (Watkins and Morgan 1955a) that the destruction of the particular group specificity by the action of these different enzymes can be specifically inhibited by the addition of the corresponding sugar to the reaction mixture.

From these types of experiment the nature of the sugars specifically concerned in the A, B and H determinants was deduced. But the detailed elucidation of the structures involved was finally dependent on the isolation and characterisation of numerous short chained oligosaccharides obtained by partial acid and alkaline hydrolysis of purified glycoproteins with different group specificities (Cote and Morgan 1956, Schiffman et al. 1962, 1964; Painter et al. 1962, Rege et al. 1964a). Determination of the structures of these fragments as well as the study of their behaviour in the inhibition and enzyme systems mentioned above eventually led to the elucidation of the oligosaccharide sequence shown in fig. 7.2.

Progress in establishing the exact nature of the group specificities of the active glycolipid materials obtained from red cells is much less advanced. However it has been shown that the carbohydrate moiety of the molecules contains the same sugars as are present in the group specific glycoproteins.

Also serological and enzymic inhibition tests indicate that here also N-acetylgalactosamine and D-galactose are the major determinant units in A and B specificity respectively (Watkins et al. 1964). So it appears possible that the same or very similar oligosaccharide groupings may be responsible for the group specificities of both the 'water soluble' glycoproteins occurring in the mucous secretions and the 'alcohol soluble' glycolipids present in the red cell membrane.

All these results suggest that the *A* and *B* alleles at the *ABO* gene locus act by determining the formation of specific glycosyl transferring enzymes which add either N-acetylgalactosaminosyl or D-galactosyl units to the ends of the carbohydrate chains in the final stages of the synthesis of these group specific macromolecules. One may suppose that whatever *A*, *B* or *O* genes are present, the synthesis of the macromolecule proceeds in much the same way as far as the formation of a substance containing multiple carbohydrate chains with H specificity (as shown in fig. 7.2) is concerned. Then in individuals carrying the *A* allele, and hence possessing the corresponding transferase enzyme, N-acetylgalactosamine residues will be added as the terminal units of the oligosaccharide chains. Similarly in individuals carrying the *B* allele, and so forming a specific D-galactosyl transferase enzyme, galactose is added as a terminal unit to the oligosaccharide chains. In individuals homozygous for the *O* allele, no corresponding enzyme is apparently present, and so no further addition to the carbohydrate chain takes place. The H specific grouping is left exposed and is therefore serologically detectable. Hence H reactivity is most marked in group O individuals.

This general hypothesis of the action of the *A*, *B* and *O* alleles is also consistent with finding that a single glycoprotein molecule may exhibit more than one blood group specificity. It has been shown for example that when a purified group specific preparation obtained from an AB subject is precipitated with an antiserum specific for A, both A and B activities are carried down in the precipitate (Morgan and Watkins 1956). Presumably in the synthesis of the glycoprotein macromolecules in group AB individuals, there is competition for the completion of the oligosaccharide chains. Any one chain may be completed either as an A active structure by the addition of N-acetylgalactosamine, or a B active structure by the addition of D-galactose. But since there are many chains on the same macromolecule some will be completed in one way and others in the other. It has also been shown that H activity may be exhibited by the same glycoprotein macromolecule which shows A or B activity (Watkins and Morgan 1957b). Evidently here, not all the chains available are completed by the addition of the A or B determinants,

and the H reactivity is due to the presence of these chains to which no additional sugar has been added.

Direct evidence for the occurrence of specific glycosyltransferase enzymes which can be regarded as the products of the *A* and the *B* alleles at the ABO locus has recently been obtained. Thus an α-D-galactosyltransferase has been shown to occur in submaxillary glands and also in gastric mucosal material, from individuals who are group B or AB, but absent in tissues from group A or O individuals. This enzyme transfers D-galactose from uridine diphosphate galactose to oligosaccharides containing at the terminal non-reducing end the H active structure α-L-fucosyl-(1→2)-galactose (Ziderman et al. 1967, Race et al. 1968). Similarly an α-N-acetyl-D-galactosaminyltransferase which transfers N-acetyl-D-galactosamine from uridine diphosphate N-acetyl-D-galactosamine to the same oligosaccharide acceptors has been demonstrated in submaxillary glands (Hearn et al. 1968) from group A or group AB individuals, but not from group B or group O individuals. Also in human milk the N-acetyl-D-galactosaminyltransferase has been found in A and AB but not in B or O individuals (Kobata et al. 1968a); and the D-galactosyltransferase has been found in milk from B and AB individuals but not in A or O individuals (Kobata et al. 1968b).

As yet little is known about the nature of the specific glycosyltransferase enzymes determined by the *A* and *B* alleles. One would expect that the different enzyme proteins are very similar in structure and perhaps only differ by a single aminoacid substitution. However this difference is presumably sufficient to result in a clear difference in substrate specificity. The 'A' enzyme presumably has a specificity for a nucleotide diphosphate compound containing an N-acetyl-D-galactosamine grouping. Similarly the 'B' enzyme has a specificity for a substrate containing a D-galactose grouping. It is noteworthy that N-acetyl-D-galactosamine and D-galactose are structurally identical except that the N-acetyl amino group at carbon No. 2 in the former, is replaced by an hydroxyl group in the latter. A relatively small difference in enzyme protein structure may well, by altering the conformation of the active site give rise to this difference in substrate specificity. The 'inactivity' of the *O* allele might be due to the formation of an enzyme protein molecule with a slightly modified structure which renders it devoid of glycosyl transferase activity. Other possibilities are a true failure in the synthesis of the enzyme protein, or the synthesis of an extremely unstable form of the enzyme protein so that its half life is very much shorter than those of the 'A' or 'B' enzymes.

7.2. The 'secretor' and the 'H' loci

In most people the saliva and other mucous secretions contain water soluble glycoproteins with A, B or H specificities according to the individual's red cell group. However there are some individuals in whom although there is no lack of glycoproteins in their secretions, the substances present are devoid of the characteristic A, B or H specificities. In fact individuals can be divided into two sharply distinct classes in this way (Lehrs 1930, Putkonen 1930). One class, the so-called 'secretors' exhibit A, B or H specificity in their saliva and other secretions, the other class called 'non-secretors' do not. Among Europeans about 80% of people are 'secretors' and about 20% 'non-secretors'.

The 'secretor' or 'non-secretor' status of an individual is constant and is genetically determined. Family studies have shown that the dimorphism is determined by a pair of allelic genes now referred to as *Se* and *se* (Schiff and Sasaki 1932). An individual carrying the *Se* allele whether homozygous *SeSe* or heterozygous *Sese* is a 'secretor'. An individual homozygous for the other allele i.e. *sese* is a 'non-secretor'. These alleles occur at a locus which is distinct from the ABO locus, and it has been shown that the two loci are not clearly linked. They either lie well separated on the same chromosome, or occur in different chromosomes.

The *Se* allele appears to be necessary for the formation of the H specific grouping in the carbohydrate chains of the water soluble glycoproteins. If it is not present, as occurs in homozygotes *sese*, then the H specific grouping is not formed. Since this appears to be a necessary prerequisite before the A or B terminal groupings can be added to the chains, A or B specificities are also absent in non-secretors even though the individual may carry the corresponding *A* or *B* alleles.

However the alleles at this so-called 'secretor' locus only affect the synthesis of 'water soluble' group specific substances. They do not appear to influence the formation of the 'alcohol soluble' substances, since A, B and H specificities are exhibited in the ordinary way by the red cells of 'non-secretors'. Furthermore the apparent block in 'non-secretors' in the formation of the characteristic carbohydrate chains of the glycoproteins, evidently only affects the last stages of their synthesis, because as will be seen later the glycoproteins actually present in the secretions of 'non-secretors' are qualitatively very similar to those in 'secretors' and only appear to lack the terminal L-fucosyl residue characteristic of the H specific chains, and the terminal N-

acetylgalactosamine or D-galactose residues which determine the A or B specificities.

There is yet another gene locus which is intimately concerned with the formation of the H specific grouping and consequently also the formation of A and B specificities. It was recognised following the discovery of certain rare individuals with a quite unusual pattern of blood group specificities (Bhende et al. 1952). The individuals in whom this peculiarity was first discovered came from Bombay, and it is consequently often referred to as the 'Bombay phenotype'. It is also called O_h. In these individuals the red cells are not agglutinated by anti-A, anti-B or anti-H, and the serum contains not only anti-A and anti-B but also anti-H in high titre. Furthermore no A, B or H specificities are detectable in saliva. The genetical basis of this unusual situation became clear when a detailed study of a family in which there occurred three individuals with this peculiar phenotypic pattern was made (Levine et al. 1955). The pedigree is shown in fig. 7.3. II_3, II_4 and II_6 are the

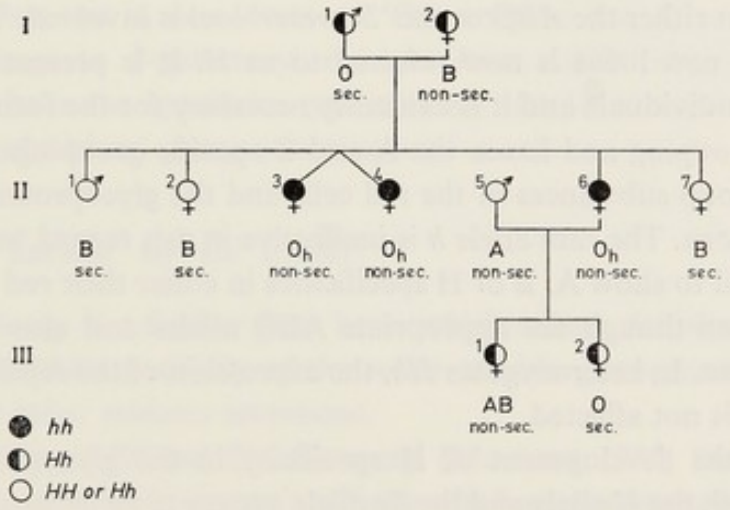


Fig. 7.3. Pedigree showing segregation of unusual blood group phenotype O_h (Bombay phenotype), and its interpretation in terms of the segregation of the rare allele *h* at the *H* locus (after Levine et al. 1955). Individuals I_1 and I_2 were first cousins. The apparent ABO types as determined by red cell agglutination tests with anti-A and anti-B are shown for each individual. Also shown are the secretor (sec.) or non-secretor (non-sec.) phenotypes as determined by standard methods on the saliva. Individuals with O_h or Bombay phenotype lack A, B and H antigens in red cells and secretions. Their sera contain the antibodies anti-A, anti-B and anti-H.

individuals with the unusual phenotype (O_h). The essential point is that one of the children of II_6 shows B reactivity in her red cells though this antigen is apparently not present in II_6 herself or her husband who is group A. Further-

more the other child of II_6 who appeared to be a normal group O individual showed H activity in her saliva and was therefore presumably a 'secretor'. Yet her father II_5 was a 'non-secretor', and her mother II_6 also showed no A, B or H activity in her saliva and might therefore also have been considered to be a 'non-secretor'. Thus it would appear that II_6 is carrying a *B* allele at the *ABO* locus though its activity is not expressed, and that she is also carrying an *Se* allele at the *Secretor* locus though neither B or H activity is present in her secretions. The simplest explanation of the findings is that II_6 and also II_3 and II_4 who show the same peculiar phenotypic pattern are homozygous for a rare allele at yet another gene locus, and that this has the effect of suppressing the formation of B and H (and probably also A) specificities in both red cells and secretions. Further evidence that homozygosity for a rare allele is involved is provided by the fact that the parents of the affected individuals were first cousins.

A number of other examples of this phenomenon has since been found, and they generally confirm the interpretation that a rare allele at a gene locus different from either the *ABO* or the '*Secretor*' loci is involved. The common allele at this new locus is now referred to as *H*. It is present in the great majority of individuals and it is evidently necessary for the formation of the H specific grouping and hence the A and B specific groupings, in both the glycolipid group substances of the red cells and the glycoprotein substances of the secretions. The rare allele *h* is ineffective in this regard, so that homozygotes *hh* fail to show A, B or H specificities in either their red cells or their secretions even though the appropriate ABO alleles and also the *Se* allele may be present. In heterozygotes *Hh*, the expression of the ABO alleles or of the *Se* allele is not affected.

Thus for the development of H specificity in the glycoproteins of the secretions both the *H* allele and the *Se* allele appear to be necessary. However in the formation of H specificity in the glycolipids of the red cell, only the *H* allele appears to be essential while *Se* seems to play no significant part.

In the biosynthesis of the blood group substances the appearance of H specificity apparently depends on the addition of L-fucose in α-(1→2) linkage to a terminal galactose residue in an oligosaccharide chain so as to give the structures shown in fig. 7.2. This step is an essential preliminary to the action of the glycosyl transferase enzymes determined by the *A* and *B* alleles of the ABO locus. Presumably it requires a specific L-fucosyl transferase capable of transferring an L-fucose residue from a suitable donor substrate to the terminal galactose residue of the oligosaccharide chain. It has been suggested (Watkins 1966) that this fucosyl transferase is the specific enzyme product of

the *H* allele and is not formed by the *h* allele. However for the enzyme to occur and also take part in the biosynthesis of the 'water soluble' group specific substances in the secretions, the allele *Se* must also be present. So 'water soluble' substances with H specificity will be formed in individuals who are either *HH* or *Hh*, and also *SeSe* or *Sese*, but not in those who are *hh* and *sese*. Since the 'Secretor' locus appears to play no part in the biosynthesis of the 'alcohol soluble' blood group substances, H specificity may be found in red cells even though the individual is a 'non-secretor' i.e. *sese*. It will not however occur in red cells and neither will the A or B specificities if the individual is of the rare 'Bombay' type, *hh*. On this view then, the '*Secretor*' locus is regarded as regulating the activity of the *H* locus, in the biosynthesis of the group specific 'water soluble' glycoproteins of the secretions, but not in the biosynthesis of the group specific 'alcohol soluble' glycolipids.

Shen et al. (1968) have studied the fucosyltransferases present in milk from secretors and non-secretors. They found that one of the enzymes was present in secretors but not in non-secretors. This enzyme transferred L-fucose in α-(1→2) linkage to β-galactosyl residues from GDP-L-fucose, and thus appeared to have the specificity predicted for the formation of H reactive groupings in the blood group substances.

7.3. *The 'Lewis' or Le locus*

The '*Lewis*' locus is a further gene locus which affects the biosynthesis and hence the antigenic specificity of the water soluble glycoproteins occurring in saliva and other mucous secretions.

Its recognition followed the discovery (Mourant 1946) that occasional individuals have in their sera an antibody which causes agglutination of red cells of some 18% of European individuals. It was subsequently found however that in as many as 90% of the population, substances which also reacted specifically with this antibody could be detected in saliva and other mucous secretions (Grubb 1951). The antigen specificity involved is now referred to as Lewis a or Le^a. Secretions or red cells which exhibit this specificity are said to be Le(a+), and those which do not Le(a−).

The presence or absence of Le^a specificity in secretions depends on a pair of alleles *Le* and *le* (Grubb 1951, Ceppellini 1955). Individuals carrying the *Le* allele either as homozygotes *LeLe* or heterozygotes *Lele* exhibit Le(a+) specificity. Individuals who are homozygous *lele* do not. The locus is separate from and apparently not closely linked to the other loci known to be in-

volved in determining the group specificities of glycoproteins in the mucous secretions (e.g. the ABO locus and the Secretor locus).

However rather unexpectedly it was found that the degree of Le^a reactivity present in secretions of individuals carrying the *Le* allele is markedly dependent on their 'secretor'–'non-secretor' status. In the secretions of 'non-secretors' Le^a activity is very much greater than in the secretions of 'secretors'. Also individuals who show Le^a activity in red cells are all 'non-secretors' (Grubb

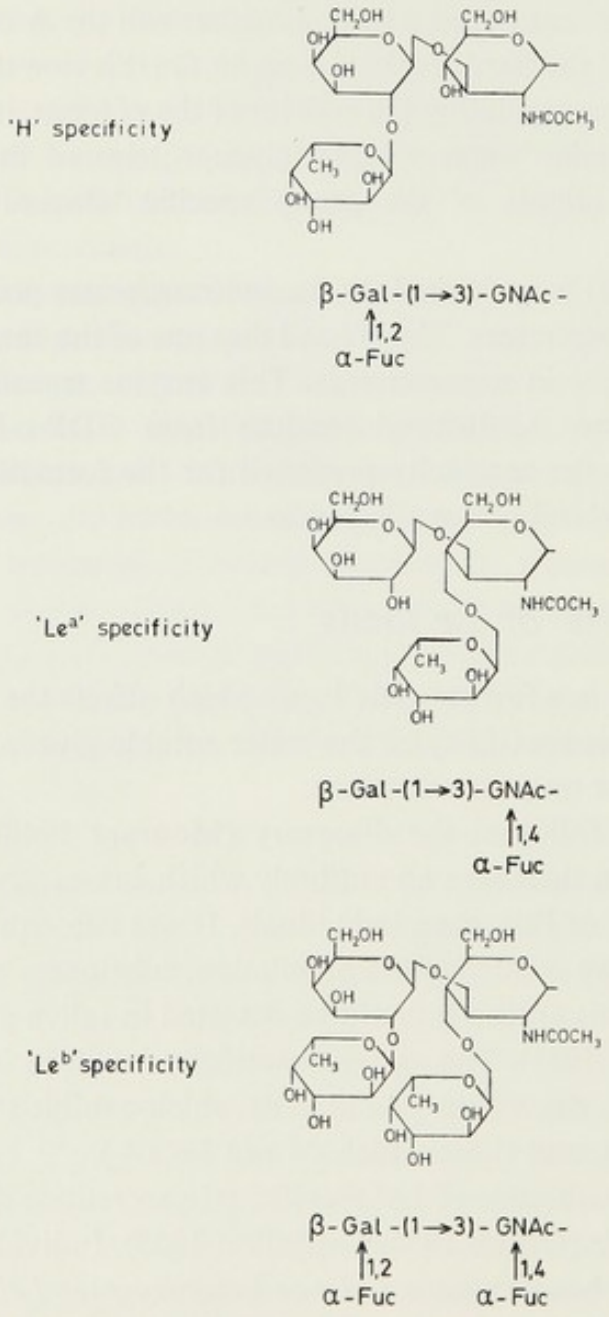


Fig. 7.4. Comparison of terminal sugar sequences in polysaccharide chains of glycoproteins which confer 'H', 'Le^a' and 'Le^b' specificities (see text). Abbreviations as in fig. 7.2.

1948, 1951). So there is an interaction between the effects of the alleles at the *Secretor* locus on the one hand and the alleles at the *Lewis* locus on the other.

The characteristic sequence of sugars which give rise to Le^a specificity in the glycoprotein molecules (Watkins and Morgan 1957a, 1962; Rege et al. 1964b) is shown in fig. 7.4. The critical sugar is the L-fucosyl residue attached in α-1,4 linkage to the penultimate N-acetylglucosaminyl residue in the chain. The main backbone of the carbohydrate chain is the same as that of one of the two types of chain which by the addition of an L-fucosyl residue to the terminal galactosyl unit can give rise to H specificity; and which by the further addition of an N-acetylgalactosaminyl or a D-galactosyl unit can give rise to A or B specificity respectively. Evidently in the biosynthesis of the glycoprotein molecules carbohydrate chains are formed which may be acted on not only by the *H*, *Se* and *A* and *B* genes but also by the *Le* gene.

In 'non-secretors' (*sese*), H reactive groupings are not formed, so that if the *Le* gene is present many carbohydrate chains are available to which Le^a specific activity can be conferred by the addition of L-fucosyl residues in the appropriate position. However in 'Secretors' (*SeSe* or *Sese*) there is presumably competition for the chains and a much smaller proportion of them are likely to end up with only the specific Le^a grouping. Others will have the H grouping and perhaps the A or B groupings according to which genes are present. Furthermore it has been shown that some chains may acquire under the influence of the *Le* gene, an L-fucosyl residue attached to the penultimate sugar residue (N-acetylglucosamine), and also under the influence of the *H* and *Se* genes an L-fucosyl residue attached to the next sugar residue (D-galactose). This compound grouping exhibits neither Le^a reactivity or H reactivity. Instead a new specificity known as Le^b and detectable only by a quite different antiserum, is produced (Marr et al. 1967). The apparent interaction between the Secretor and the Lewis loci, whereby much greater Le^a activity is found in 'non-secretors' (*sese*) who carry the *Le* gene, than in 'secretors', who carry the *Le* gene, is thus readily accounted for.

It seems likely that the *Le* gene does not play a specific role in the biosynthesis of the glycolipid group specific substances of the red cell, since it has been shown that red cell Le(a+) reactivity which is normally exhibited only by individuals who are non-secretors (*sese*) is largely if not entirely due to adsorption on the red cell of Le^a active substances formed elsewhere but which occur in relatively greater amounts in 'non-secretors' (Sneath and Sneath 1955, Marcus and Cass 1969).

7.4. The biosynthetic pathways for the group specific glycoproteins

According to the behaviour of the red cells and the mucous secretions of different individuals with the five specific antibodies, anti-A, anti-B, anti-H, anti-Le^a and anti-Le^b it is possible to define six distinctive types of reaction as shown in table 7.2.

TABLE 7.2

Six types of individuals distinguishable on the basis of the reactions of their red cells and secretions with the antibodies anti-A, anti-B, anti-H, anti-Le^a and anti-Le^b (Watkins 1966).

Gene combination			Specificities detectable on red cells			Specificities detectable in secretions			
H locus	*Se* locus	*Le* locus	ABH	Le^a	Le^b	ABH	Le^a	Le^b	
1. *HH* or *Hh*	*SeSe* or *Sese*	*LeLe* or *Lele*	+++	—	++	+++	+	++	
2. *HH* or *Hh*	*sese*	*LeLe* or *Lele*	+++	+++	—	—	+++	—	
3. *HH* or *Hh*	*SeSe* or *Sese*	*lele*	+++	—	—	+++	—	—	
4. *HH* or *Hh*	*sese*	*lele*	+++	—	—	—	—	—	
5. *hh*	*SeSe* or *Sese*	*LeLe* or *Lele*	—	+++	—	—	+++	—	'Bombay' types
6. *hh*	*sese*	*lele*	—	—	—	—	—	—	'Bombay' types

Strong specific activity +++; weak specific activity +; no activity —.

The fourth set of reactions shown in table 7.2 is of special interest because in these individuals although the red cells have the standard reactions with anti-A, anti-B, and anti-H according to the particular ABO genotype of the individuals, the mucous secretions fail to react with any of the antibodies. These individuals constitute about 2% of European populations and are homozygous for the 'inactive' alleles at both the '*Secretor*' and the '*Lewis*' loci. That is they have the genotype *sese*, *lele*. The secretions of such individuals are however not devoid of glycoproteins of the kind which in other people exhibit the different group specificities and in fact such glycoproteins have been isolated and studied. They can be regarded as precursors of the other group specific substances, in which the biosynthesis has been stopped before groupings giving rise to the different characteristic specificities have been added to the carbohydrate chains (Watkins 1966).

This so called 'precursor' substance appears to have at least two sorts of carbohydrate chains (Rege et al. 1964a) and these may be joined through a branch point to a common chain linking to the polypeptide backbone (Lloyd et al. 1966, 1968). Both these chains appear to have a terminal galactosyl residue at the non reducing end (table 7.3). This is joined to an N-acetylglucosaminyl residue in one chain (type 1) by a β-1,3 linkage, and in the other chain (type 2) by a β-1,4 linkage. The substance has been shown to react with antisera prepared against Type XIV pneumococcus, and this specificity can be attributed to the terminal structure of the type 2 chain.

A variety of gene loci presumably play a part in determining the biosynthesis of this so-called 'precursor' glycoprotein. They would be concerned in determining the synthesis of the various enzymes necessary for assembling the carbohydrate chains. And no doubt at least one further locus is necessary to define the sequence of aminoacids in the polypeptide backbone to which the carbohydrate chains are thought to be attached. However nothing is known about these loci because so far no genetical differences in the 'precursor' substance as it occurs in different individuals have been detected. The four loci *ABO*, *Secretor*, *H* and *Lewis*, about which something is known, are evidently concerned only with the later stages of the biosynthesis of their glycoproteins and appear to act only after synthesis as far as the 'precursor' has occurred.

Table 7.3 (Morgan and Watkins 1969) summarises the way in which the various genes at these several loci are thought to act in the formation of group specific glycoproteins of mucous secretions. It is supposed that if an individual carries the *Le* gene, he forms a specific fucosyl transferase enzyme which is capable of adding an L-fucose unit to the carbon No. 4 position of the subterminal N-acetylglucosamine unit in the type 1 chains of the 'precursor'. The type 2 chains are unaffected because in the corresponding N-acetylglucosamine unit the carbon no. 4 position is already substituted. If the individual carries the *H* allele at the '*H*' locus and also the *Se* allele at the '*Secretor*' locus gene, then he forms a different fucosyl transferase enzyme which is capable of adding an L-fucose unit to the carbon no. 2 position of the terminal galactose unit in either the type 1 or the type 2 chains. The prior addition of this fucose unit to form the H type structure appears to be necessary before the specific transferases thought to be determined by the *A* and *B* alleles at the ABO locus, can act. The 'A' transferase adds an N-acetylgalactosamine unit in α-linkage to the carbon no. 3 of the terminal galactose units of the H active structures. The 'B' transferase adds D-galactose in α-linkage to the same position. The alternative alleles at the

TABLE 7.3

Genetic control of the biosynthesis of structures with the blood group specificities A, B, H, Le^a, and Le^b (Morgan and Watkins 1969). Abbreviations as in fig. 7.2. The numerals in brackets refer to two different L-fucosyl transferases referred to in the text (pp. 202-203 and 205, 207).

Gene	Enzyme product	Chain-ending in precursor	Structure	Serological specificity
—	—	Type 1	β-Gal-(1→3)-GNAc--	—
—	—	Type 2	β-Gal-(1→4)-GNAc--	Type XIV
H	α-L-Fucosyltransferase (1)	Type 1	β-Gal-(1→3)-GNAc-- ↑ 1, 2 α-Fuc	H
		Type 2	β-Gal-(1→4)-GNAc-- ↑ 1, 2 α-Fuc	H
Le	α-L-Fucosyltransferase (2)	Type 1	β-Gal-(1→3)-GNAc-- ↑ 1, 4 α-Fuc	Le^a
		Type 2	β-Gal-(1→4)-GNAc--	Type XIV
H and *Le*	α-L-Fucosyltransferases (1) and (2)	Type 1	β-Gal-(1→3)-GNAc-- ↑ 1, 2 ↑ 1, 4 α-Fuc α-Fuc	Le^b
		Type 2	β-Gal-(1→4)-GNAc-- ↑ 1, 2 α-Fuc	H
H and *A*	α-L-Fucosyltransferase (1) and α-*N*-acetylgalactosaminyltransferase	Type 1	α-GalNAc-(1→3)-β-Gal-(1→3)-GNAc-- ↑ 1, 2 α-Fuc	A
		Type 2	α-GalNAc-(1→3)-β-Gal-(1→4)-GNAc-- ↑ 1, 2 α-Fuc	A
H and *B*	α-L-Fucosyltransferase (1) and α-D-galactosyltransferase	Type 1	α-Gal-(1→3)-β-Gal-(1→3)-GNAc-- ↑ 1, 2 α-Fuc	B
		Type 2	α-Gal-(1→3)-β-Gal-(1→4)-GNAc-- ↑ 1, 2 α-Fuc	B

various loci, for example the *h* allele at the *H* locus, the *le* allele at the '*Lewis*' locus and the *O* allele at the *ABO* locus are ineffective in the biosynthetic process because they fail to lead to formation of the appropriate transferases.

Thus from the results obtained in extensive and very diverse serological, genetical and biochemical studies it has become possible to see how the group specificities of these complex glycoproteins are built up stepwise by the action of a series of glycosyl transferase enzymes determined by genes at several different loci. It is of some interest to note that while the action of a particular enzyme in the course of the biosynthesis of glycoproteins, may be adding the appropriate sugar unit to an oligosaccharide chain to produce a new antigenic specificity, it may at the same time abolish or cancel a specificity previously present. For example, the addition of the N-acetylgalactosamine unit to the H reactive structure leads to the appearance of A reactivity, but the loss of H reactivity.

It is also of interest that the sequential formation of the various group specificities can be effectively reversed experimentally by the stepwise degradation of the glycoproteins using specific enzymes which split off sugar units one at a time. This is illustrated by experiments in which purified substances exhibiting only A or B specificities were treated successively with

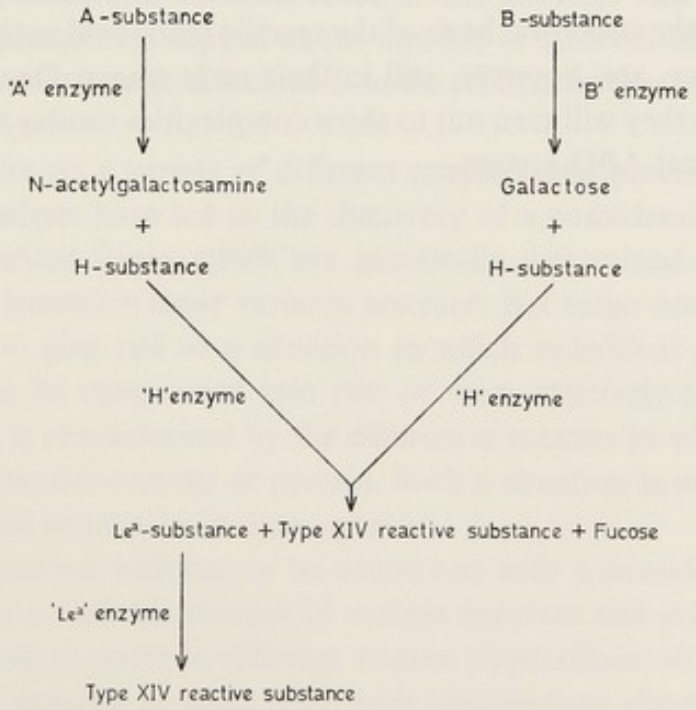


Fig. 7.5. Specificities revealed by the sequential enzymic degradation of A and B specific substances (Watkins 1966).

different enzymes capable of splitting off the characteristic sugar units associated with A, B, H and Le^a specificities (fig. 7.5, Watkins 1966). The so-called 'A' and 'B' destroying enzymes lead to a loss of A and B reactivities and the appearance of H reactivity. Then the 'H' destroying enzyme leads to the loss of the H reactivity and the appearance of Le^a and also of pneumococcal type XIV reactivity. Finally treatment with the Le^a destroying enzyme resulted in the loss of Le^a reactivity and the further enhancement of pneumococcal type XIV reactivity. In effect the various specific groupings had been stripped off leaving the so-called 'precursor' glycoprotein.

The role of the *ABO* and the *H* loci is probably much the same in the biosynthesis of the glycolipid 'alcohol soluble' group specific substances of the red cell as in the biosynthesis of the glycoprotein 'water soluble' group specific substances of the mucous secretions. However the '*Lewis*' and '*Secretor*' loci do not appear to play any part in the biosynthesis of the glycolipid substances. Exactly why this should be so, is not at present understood. The reason will perhaps emerge as the structures of the group specific glycolipid substances are elucidated.

A series of other gene loci (*Rh*, *MNS*, *Xg*, etc.) at which alleles determining inherited differences in antigenic substances which form part of the structure of red blood cells have also been identified by serological methods (Race and Sanger 1968). The study of the nature of the substances concerned, and the elucidation of the structural basis of the specific individual antigenic differences which occur, are, however, still in their early stages. One may perhaps anticipate that they will turn out to show complexities similar to those found with the classical ABO system.

CHAPTER 8

Enzyme and protein diversity in human populations

8.1. 'Common' and 'rare' variants

A very large number of different enzymes and other proteins are synthesised in the human organism and we may presume that the primary aminoacid sequence of each of their distinctive polypeptide chains is coded in the DNA of a separate gene locus. So there must be a vast array of so-called 'structural' gene loci in the genetic constitution of every individual. Furthermore as a result of mutational events which have occurred in earlier generations there may occur at any given gene locus a series of different alleles each of which determines a structurally distinct version of the particular polypeptide chain. Thus the extent to which individual members of a population differ from one another in the structural characteristics of the enzymes and proteins they synthesise, will in general depend on the number of different alleles which are present at these various gene loci, and on the relative frequencies with which they occur.

In fact studies on a variety of different enzymes and proteins in different human populations have led to the discovery of a considerable number of structurally variant forms which are genetically determined. Many of the alleles which determine these variants are rare. But some occur sufficiently frequently as to give rise to a situation in which individual members of a population can be categorised into two or more relatively common types each of which is characterised by the distinctive manner in which they synthesise the particular enzyme or protein. Such a situation is usually referred to as an enzyme or protein 'polymorphism'.

In this chapter we will mainly be concerned with a consideration of the general incidence and distribution of variant enzymes and proteins, as they have been found to occur in different human populations; with the kind of inferences it is possible to make from such observations about the extent of enzyme and protein diversity in general; and with the question as to how the actual distributions of enzyme and protein variants that we observe may have

been brought about. A general idea of the nature and dimensions of these problems can perhaps best be obtained by considering the main features of some well-studied, though rather different examples.

8.1.1. Haemoglobin

Haemoglobin A which constitutes most of the protein content of normal red blood cells, has been studied more extensively than any other human protein. It will be recalled that the protein is a tetramer and contains two distinct types of polypeptide chain (α and β) each of which is represented twice in the molecule. The α-chain has a sequence of 141 aminoacids, and is presumably coded by a gene containing a stretch of DNA with at least 423 bases, and the β-chain has 146 aminoacids presumably coded by a DNA sequence of 438 bases. The two gene loci are apparently well separated on the chromosomes and not closely linked.

As a result of extensive studies on patients with different sorts of haematological disease, and from surveys of randomly selected individuals in a variety of populations, a very large number of inherited variants of Hb A have now been discovered, and in many cases the distinctive structural peculiarity has been identified. The great majority of these variants appear to differ from the normal form, $\alpha_2\beta_2$, by a single aminoacid substitution in either the α or the β polypeptide chains, and they can be attributed to alleles at either the α or β gene loci, which originally arose by mutational events involving no more than the replacement of one base by another at a particular point in the DNA sequence of the gene (pp. 12–16). In a few cases the aberrant polypeptide chain has been found to lack one or a consecutive series of several aminoacids from its full sequence, and the alleles concerned here must presumably have arisen from mutational events which led to a deletion of a small part of the DNA sequence (pp. 86–87). There are also examples of variants (e.g. Hb Lepore) where a more extensive alteration of the DNA structure involving two adjacent genes has occurred (pp. 80–83).

Some of these haemoglobin variants give rise to quite severe haematological disease even in heterozygotes. But many appear to be relatively harmless and these are probably underrepresented in the total data because of the manner in which much of it has been collected, since patients with haematological abnormalities have been particularly selected for study. Furthermore although the search for haemoglobin variants has been extensive, it is clear that only a minute fraction of the total world population has been examined. Also the electrophoretic methods mainly used for the detection of variants are known to be capable of picking up only a proportion (perhaps about one

third) of all the variants that might occur. So although a considerable number of different variants have already been identified, it seems almost certain that these represent only a small proportion of all the structurally different variants of Hb A which actually occur among individual members of the species. Certainly judging from the current literature the rate of discovery of new examples does not seem to be slowing down.

Thus at both the α and the β gene loci which code for the two characteristic polypeptide chains of Hb A, a very large number of different alleles, generated by mutational events in the remote or more recent past, exist in human populations. But these alleles vary greatly in the relative frequencies with which they occur, and in fact it is possible to classify those that have already been discovered fairly sharply into two main groups on this basis.

The first and much the smallest group comprises those alleles which occur with a relatively high incidence in one or another human population (fig. 8.1).

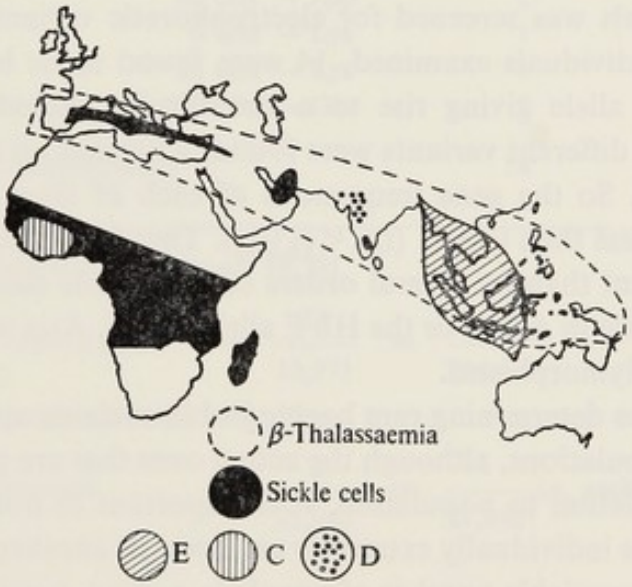


Fig. 8.1. World distribution of the major haemoglobin abnormalities. (From Lehmann et al. 1966.)

In the case of the β locus it includes the allele that determines Hb S (or sickle-cell haemoglobin) which is very common in tropical Africa where in different populations it may occur in 20% or more of all individuals; the allele determining Hb C which is more localised to W. Africa; the allele determining Hb E which is common in many populations in S.E. Asia, and the allele determining Hb D Punjab which occurs in appreciable frequency in certain populations in India (for detailed references see Livingstone 1967). Thus in

each of these populations, besides the so-called normal allele at the β gene locus, at least one other allele determining a structurally distinct form of the β polypeptide chain is relatively common, so that the individual members of the population can be classified into two or more not uncommon types according to the characteristics of the haemoglobin they synthesise. These are typical examples of what is meant by genetically determined protein 'polymorphism'.

The second group of alleles determining haemoglobin variants are in contrast all extremely rare. They have turned up in an irregular manner in a wide variety of different populations, and many have so far been seen only among individual members of a single family. An idea of the incidence of these rare alleles can be obtained from the results of some surveys of population samples in England and Denmark (Huntsman et al. 1963, Liddell et al. 1964, Sick et al. 1967, and Lehmann, personal communication). The haemoglobin from large numbers of essentially randomly selected and unrelated individuals was screened for electrophoretic variants. In all out of nearly 11,000 individuals examined, 14 were found to be heterozygous for one or another allele giving rise to a structurally altered α- or β-chain. Among them 10 different variants were found, 4 turning up twice and 6 only once (table 8.1). So the gene frequencies of each of these different alleles appeared to be less than 0.0001 (i.e. $< \frac{1}{10,000}$). Thus individually they appear to be less frequent than by several orders of magnitude than such alleles as the sickle cell allele in Africa or the Hb E allele in S.E. Asia which give rise to the so-called polymorphisms.

Multiple alleles determining rare haemoglobin variants appear to occur in most human populations, although the actual ones that are present probably vary from population to population. It is important to note that although these variants are individually extremely rare, one or another of them may be present in an appreciable number of people in a given population. Thus the results of Lehmann and his colleagues suggest that rather more than 1 in 1,000 individuals in Europe are heterozygous for an allele determining an electrophoretic variant of Hb A. One may reasonably predict that these represent only about one third of the structural variants due to single amino-acid substitutions that actually occur, the others not being detectable by electrophoresis. So it is quite likely that in Northern Europe as many as 1 in 300 people are heterozygous for an allele causing a structural variant of Hb A.

The high incidence of the specific alleles which give rise to the so-called haemoglobin polymorphisms in certain parts of the world, has generally been explained by the hypothesis that heterozygotes for such alleles are at

TABLE 8.1

Rare haemoglobin variants found in electrophoretic surveys in English and Danish populations (data from Huntsman et al. 1963, Liddell et al. 1964, Sick et al. 1967, and Lehmann, personal communication 1969).

Affected polypeptide chain	Aminoacid substitution	Number of heterozygotes in total sample
α	15 Gly → Asp	2
α	53 Gly → Asp	1
α	47 Asp → His	1
β	16 Gly → Asp	2
β	43 Glu → Ala	1
β	47 Asp → Asn	1
β	69 Gly → Asp	2
β	6 Glu → Lys	1
β	26 Glu → Lys	1
β	121 Glu → Gln	2
Total heterozygotes		14

Total number of individuals examined: 10,791.

Total incidence of heterozygotes $= \frac{14}{10{,}971} = \frac{1}{784}$

Number of different rare variants $= 10$

Separate allele frequencies $= \frac{2}{21{,}942}$ or $\frac{1}{21{,}942}$ i.e. each <0.0001

some selective advantage compared with normal homozygotes under the environmental conditions that prevail or have prevailed in the past in these areas. One may suppose that the primary aminoacid sequences of a protein such as haemoglobin have been evolved by natural selection and have come to be more or less optimal for the species, so that mutant alleles which alter the structure of the protein are unlikely in the great majority of cases to confer any biological advantage on individuals who carry them, and many will be in some degree deleterious. But very occasionally a mutant may appear which results in a positive selective advantage in the sense that individuals carrying the allele contribute on average more offspring to the next generation

than normal homozygotes, at least under certain environmental conditions. Such an allele will, in general, tend to increase in frequency in successive generations providing the particular environmental conditions which favour it persist. However, if the specific advantage is peculiar to the heterozygous state, so that the heterozygote contributes more to the next generation than either type of homozygote, an equilibrium situation may eventually be established in which the relatively greater contribution of the variant allele to the next generation by the heterozygote is in effect balanced by the relatively reduced contribution from the homozygote. Such a situation is usually referred to as a balanced polymorphism.

The sickle-cell polymorphism in Africa provides the clearest example of this general type of phenomenon (Allison 1954, 1964). Homozygotes with sickle-cell disease tend to die in early life and very few survive to adult life and have children. So the particular sickle-cell alleles carried by homozygotes tend to be selectively eliminated from the population in each successive generation. Yet the sickle-cell allele is extremely common in many populations living in Africa, and since the possibility of a special and exceptionally high rate of mutation can be excluded (Vandepitte et al. 1955), the remarkable incidence of this particular allele in these populations could only have come

TABLE 8.2

Incidence of sickle-cell trait among African children whose deaths could be attributed to malaria (Allison 1964).

Locality	Deaths due to malaria	Number with sickle-cell trait	Incidence of sickle-cell trait in the population	Expected number with sickle-cell trait if no selective differential	References
Uganda (Kampala)	16	0	0.16	2.6	Raper (1956)
Congo (Leopoldville)	23	0	0.235	5.4	Lambotte-Legrand, J. and C. (1958)
Congo (Luluaborg)	23	1	0.25	5.7	Vandepitte (1959)
Ghana (Accra)	13	0	0.18	2.3	Edington and Watson-Williams (1964)
Nigeria (Abadan)	29	0	0.24	7.0	Edington and Watson-Williams (1964)
Totals	104	1		23.0	

about and subsequently been maintained in the face of the marked selective pressure against the sickle-cell homozygotes, if the heterozygotes have enjoyed a significant selective advantage over the normal homozygotes.

There is indeed a not inconsiderable body of evidence which indicates that such a selective advantage in fact exists for sickle-cell heterozygotes in Africa, and is due to their better chance of surviving to adult life in areas where malaria (*P. falciparum*) is an important cause of morbidity and mortality. Table 8.2, for example, summarises data collected in different places on the incidence of sickle-cell heterozygotes in children dying from malaria. Among the 104 deaths due to malaria, only one occurred in a sickle-cell heterozygote although about 23 might have been expected if mortality from malaria had been the same in sickle-cell heterozygotes as in other members of these populations. Data from morbidity surveys point in the same direction. Table 8.3 shows the results of a study in which all the patients admitted to a children's ward in Kampala (Uganda) in a given period were classified

TABLE 8.3

Incidence of sickle-cell trait amongst 818 consecutive admissions to a children's ward at Kampala (thirty-one patients with sickle-cell anaemia admitted during this period are not included). (After Raper 1956.)

Disease group	Total	Number with sickle-cell trait	Incidence of sickle-cell trait
Miscellaneous	186	25	0.13
Pneumonia	118	18	0.15
Upper respiratory infections	59	13	0.22
Diarrhoea and vomiting	106	25	0.24
Poliomyelitis	26	4	0.15
Tuberculosis	37	8	0.22
Meningitis (purulent)	26	5	0.19
Malnutrition	77	11	0.14
Hookworm anaemia	30	2	0.07
Typhoid fever	17	6	0.35
Malaria (a) Uncomplicated	83	13	0.16
(b) Cerebral	47	—	0.00
(c) Blackwater fever	6	—	0.00
Total admissions	818	130	0.16

according to whether or not they had the sickle-cell trait (Raper 1956). The outstanding finding is the absence of any sickle-cell heterozygotes among the patients admitted with cerebral malaria. In patients with uncomplicated malaria however, the incidence of the sickle-cell trait was not noticeably different from that in patients with other conditions. Similar data have also been obtained elsewhere, and they indicate that sickle-cell heterozygotes are significantly less liable to the development of cerebral malaria.

Other evidence indicating that sickle-cell heterozygotes, particularly in childhood, are less severely affected by malaria has come from comparisons of the degree of *P. falciparum* parasitaemia in sickle-cell heterozygotes and in other individuals in areas where the sickle-cell allele is common and malaria is endemic (for detailed review see Allison 1964). It has also been shown that in general a high frequency of the sickle-cell allele is only found in populations living in regions where malaria is, or was until recently, endemic, or in population groups such as the Negroes in the U.S.A. whose ancestors came from such regions.

In general it appears that the malarial parasite, *P. falciparum*, flourishes less well in individuals who are heterozygous for the sickle-cell allele, and who have a mixture of both Hb A and Hb S in their red cells, than in individuals whose red cells contain only Hb A. This evidently results in sickle-cell heterozygotes having a reduced susceptibility in childhood to the development of the more serious complications of malaria and to a reduced mortality, so that where malaria is a major cause of death in early life, sickle-cell heterozygotes have a better chance of surviving to become adults and so to contribute offspring to the next generation. This differential susceptibility to the effects of malaria appears to be most marked in early life. Later on it tends to be obscured by the development of acquired immunity.

The precise biochemical reasons for this differential susceptibility to the effects of malaria are still rather uncertain. Also it is difficult to evaluate quantitatively in any particular population the degree of selective advantage enjoyed by the sickle-cell heterozygotes. However, it is possible to estimate theoretically the selective differential that would be required to maintain an allele at a given frequency in a population by the selective survival of heterozygotes. A simple model (Penrose 1954, 1963) of a population in equilibrium with balanced polymorphism is shown in table 8.4. It will be seen that if homozygotes (*aa*) for a particular allele always die in early life so that they contribute nothing to the next generation, the allele may nevertheless be maintained with a frequency as high as 0.1 if the heterozygotes (*Aa*) have a selective advantage of about 11% over the 'normal' homozygotes (*AA*).

TABLE 8.4

Population model illustrating balanced polymorphism due to selective advantage of the heterozygote (Penrose 1954, 1963).

Genotypes	Relative frequencies at conception (zygotic genotypes)	Relative fitness	Relative frequencies among parents (parental genotypes)	Example assuming $p = 0.9$, $q = 0.1$, and genotype *aa* is lethal in early life		
	(i)	(ii)	(iii) = (i)×(ii)	(i)	(ii)	(iii)
AA	p^2	$1-k/p^2$	$p^2(1-k/p^2)$	81	0.988	80
Aa	$2pq$	$1+k/pq$	$2pq(1+k/pq)$	18	1.111	20
aa	q^2	$1-k/q^2$	$q^2(1-k/q^2)$	1	0.000	0
All types	1	—	1	100	—	100

The three genotypes *AA*, *Aa* and *aa* differ in their relative fitnesses, that is, their chances of becoming parents of the next generation. However, the frequencies p and q of the two alleles *A* and *a* remain the same from one generation to the next. Random mating is assumed.

The example shows the relative fitnesses of individuals of genotypes *AA* and *Aa* which would be required to maintain a balanced polymorphism with $p = 0.9$ and $q = 0.1$, assuming that individuals of genotype *aa* die in early life (i.e. fitness = 0). The incidence of the heterozygote among the parents is 20%.

In such circumstances some 20% of the adult population would be heterozygotes. This kind of model makes it plausible to suppose that the observed differences in morbidity and mortality from malaria in different parts of Africa are, or have been in the past, sufficient to produce a selection differential of the required order of magnitude.

The polymorphisms of other haemoglobin variants, such as Hb C in W. Africa and Hb E in S.E. Asia, have also been attributed to a selective advantage of the heterozygotes due to differential susceptibility to malaria. However, although the geographical distribution of these polymorphisms is very suggestive, there is as yet little direct evidence to support the idea. The same is true for the thalassaemia polymorphisms, where again the likely importance of malaria as a selective agent is indicated by the geographical distributions of the alleles. Thus in Italy (Bianco et al. 1952) a relatively high frequency (0.05–0.10) of one or other of the β-thalassaemia alleles is partic-

ularly marked in the Ferrara area, Sardinia and the South, and it is in these parts of the country that endemic malaria is, or has been until recently, a particularly important cause of mortality. Similar correlations between the distributions of thalassaemia and malaria are also seen elsewhere. One may note that Cooley's disease or 'thalassaemia major' which occurs in individuals homozygous for these alleles is in general a very severe condition and considerably reduces the chances of survival of affected subjects to adult life, so that for the thalassaemia alleles which cause this condition to have reached the high frequency observed in many areas, it seems necessary to postulate a quite significant selective advantage for the heterozygotes.

8.1.2. Glucose-6-dehydrogenase (G-6-PD)

Nearly thirty different variant forms of G-6-PD have now been discovered (pp. 121–131). They differ from the normal form of the enzyme and from one another in such properties as electrophoretic mobility, Michaelis constants, thermostability and *p*H optima, and it seems very probable that most or all of them are due to single aminoacid substitutions in the protein, similar to those found in the haemoglobins. They are apparently determined by a series of alleles at a gene locus on the X chromosome.

The incidence and distribution of these alleles in different populations is in a number of respects similar to that found with the haemoglobin alleles. Many are evidently rare, but some have an unusually high incidence in particular populations and give rise to characteristic polymorphisms. For instance, besides the allele Gd^{B} which determines the normal form of the enzyme, two other alleles, Gd^{A-} and Gd^{A}, both occur in many African populations with gene frequencies of around 0.2, though they are rare or absent elsewhere. The variant protein determined by Gd^{A-} causes the well-known Negro form of G-6-PD deficiency which is the basis of primaquine sensitivity and certain other adverse drug reactions. However, apart from this drug idiosyncrasy individuals carrying this allele appear to be in other respects quite healthy. The other common variant in Negroes determined by the allele Gd^{A} is associated with only a very slight reduction in enzyme level, and this apparently is harmless.

In many populations living in Southern Europe and the Middle East, a different sort of G-6-PD polymorphism occurs due to the high incidence of the allele $Gd^{Mediterranean}$. This determines another striking form of G-6-PD deficiency, and it predisposes to the haemolytic disease known as favism which may occur when affected individuals eat fava beans, a common feature of the diet in this part of the world. There are probably also other

G-6-PD alleles which occur commonly in particular areas, for example Gd^{Canton} in S.E. Asia and Gd^{Athens} in Greece, although their distributions have not yet been worked out in detail.

Because populations which have a high incidence of one or another form of G-6-PD deficiency come from areas in which malaria is or has been in the past a major cause of mortality, it has been suggested (Motulsky 1964) that here, as in the case of the sickle-cell gene, malaria may have been an important selective agent in determining the prevalence of particular G-6-PD alleles (e.g. Gd^{A-} in Negro populations and $Gd^{Mediterranean}$ in Southern European and Middle Eastern populations). The malaria parasite might proliferate less well in individuals whose red cells were G-6-PD deficient and therefore in some degree metabolically abnormal. There is some, though as yet not very extensive evidence (Gilles et al. 1967, Luzzatto et al. 1969) to suggest that these alleles may indeed confer some selective advantage in terms of malarial morbidity or mortality. But one must also note that the Gd^{A} allele which, though as prevalent as the Gd^{A-} allele in Africa and similarly rare or absent elsewhere, does not result like the Gd^{A-} in a marked enzyme deficiency, and would be expected to alter the metabolism of the red cell hardly at all.

8.1.3. The haptoglobin variants

The various haptoglobin types (ch. 3, pp. 65–75) provide another example of protein variation which has been studied very extensively. There are two sorts of polypeptide chains, α and β, and most of the variants which have been observed can be attributed to alleles at the α gene locus. The findings here however are in striking contrast to those obtained with haemoglobin and G-6-PD because there are at least three alleles (Hp^{1S}, Hp^{1F} and Hp^{2}) which are common and widespread throughout the world (Kirk 1968). In European and African populations all three are found, though with differing frequencies, and in Asiatic populations Hp^{1S} and Hp^{2} both have a significant incidence, though Hp^{1F} may be rare (Shim and Bearn 1964). As with haemoglobin and G-6-PD a number of other very rare alleles at the haptoglobin loci (both α and β) have been shown to occur.

A point of special interest about the haptoglobin polymorphism is that it is possible to infer from the structural differences in the protein something about the origin of the alleles (Smithies et al. 1962b). The α polypeptides determined by Hp^{1F} and Hp^{1S} each contain 83 aminoacids, and differ only in a single one (Black and Dixon 1968). The α polypeptide determined by Hp^{2} is nearly twice as long (142 aminoacids) and appears to represent an end to end fusion of the hp1Fα and the hp1Sα polypeptides with a sequence of 24

residues missing at the site of fusion. It presumably originated as the result of a mutational event involving a chromosomal rearrangement in an individual who happened to be heterozygous for Hp^{1F} and Hp^{1S}. In other words, the new allele probably arose in a population already polymorphic for the Hp^{1F} and Hp^{1S} alleles. Furthermore, the peculiar structure of the polypeptide results in a rather characteristic polymerisation of the haptoglobin molecule which is readily detected by starch gel electrophoresis, and since this effect has not been seen in haptoglobins in other species including higher apes (Parker and Bearn 1961), it seems quite likely that the mutational event giving rise to the Hp^2 allele occurred only after the separation of the human line. Nevertheless it has apparently spread throughout the species and today is the commonest of the three alleles in most human populations (fig. 8.2).

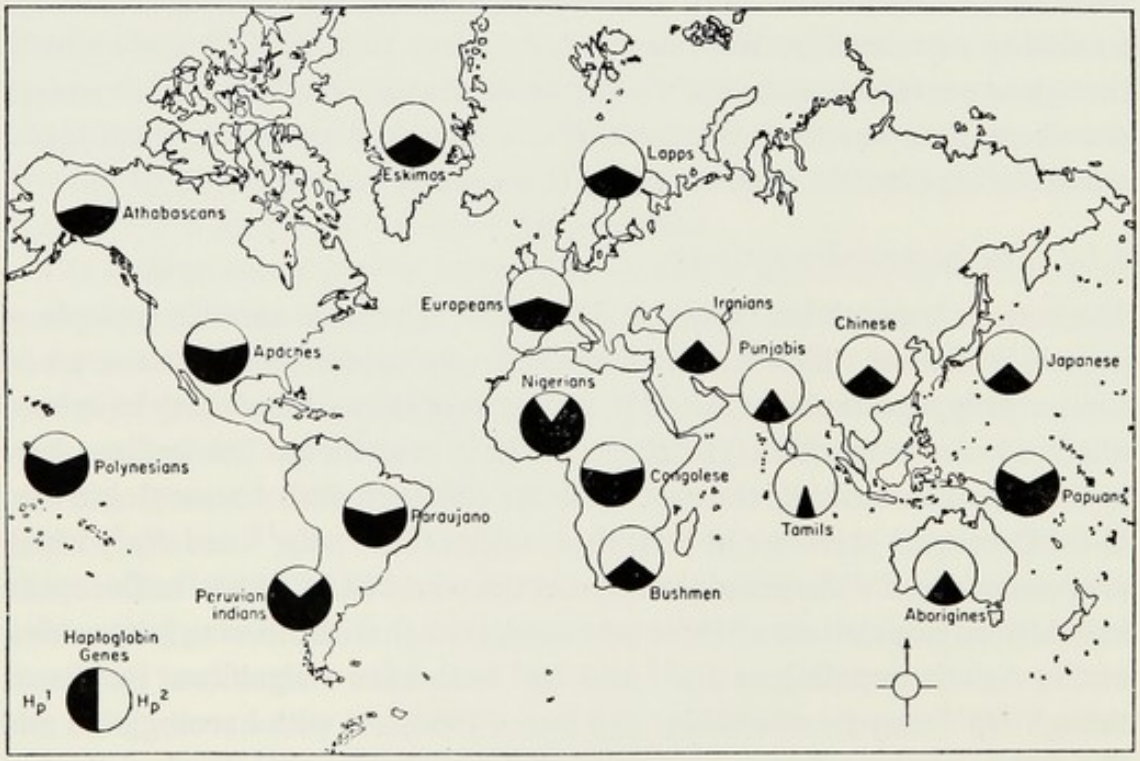


Fig. 8.2. World distribution of the Hp^1 and Hp^2 alleles (Kirk 1968).

Thus one appears to be observing an evolutionary change in the gross structure of a specific protein, and one might anticipate that the new structure conferred some distinctive selective advantage. Yet it is difficult to see from what is known about the differences in the properties of the common haptoglobin types exactly what this might be.

Unfortunately the exact significance of haptoglobin in normal function is

still very unclear. It has the characteristic property of binding tightly and specifically with free haemoglobin, a reaction which might have some significance in connexion with the conservation of iron in the body; and it has been shown (Nyman 1959) that the haptoglobin content of serum as measured by its haemoglobin binding capacity differs on average between the three types, Hp 1-1, Hp 2-1 and Hp 2-2, roughly in the ratio 135:110:85, although there is much overlap between the distributions (fig. 8.3). Also the haptoglobin–haemoglobin complex appears to be a substrate, for the liver enzyme

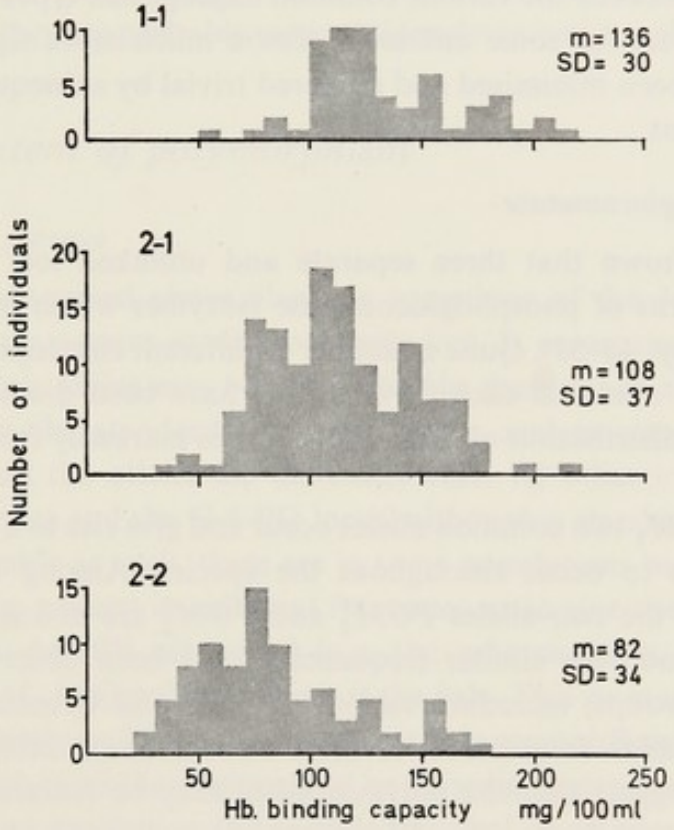


Fig. 8.3. Distributions of haemoglobin binding capacities of sera from individuals of the common haptoglobin types Hp 1-1, Hp 2-1 and Hp 2-2. The mean values (m) and standard deviations (SD) for each type are also shown. (Data of Nyman 1959.)

α-methenyl oxygenase which is thought to be concerned in bile pigment formation (Nakajima et al. 1963), and it has been reported that the Hp 2-2 haptoglobin type is somewhat more effective in this respect than Hp 2-1 or Hp 1-1 (Nakajima 1963). But these effects have not yet been shown to lead to any obvious metabolic differences, and as far as is known, individuals of the different haptoglobin types are all equally healthy. It may also be noted that

there appear to be certain as yet poorly defined haptoglobin alleles which apparently lead to a gross reduction and, in some cases, virtual absence of serum haptoglobin, and at least one of these alleles is apparently quite common in Negro populations (Giblett and Steinberg 1960), though not elsewhere. However, individuals in whom such a genetically determined deficiency of haptoglobin has been demonstrated do not appear to exhibit any other obvious phenotypic peculiarities which might be selectively significant.

In general then, it seems necessary to assume that such selective differences as may occur between the various common haptoglobin types are very slight, or that they were for some unknown reason much more significant in the past but have been minimised and rendered trivial by subsequent changes in the environment.

8.1.4. Phosphoglucomutase

It has been shown that three separate and unlinked loci determine the complex patterns of phosphoglucomutase isozymes which occur in human tissues (ch. 2, pp. 46–53). Quite a number of different electrophoretic variants determined by alleles at each of these loci have been discovered. But the incidence and distribution of these alleles varies markedly from one locus to another.

At locus PGM_1 two common alleles occur and give rise to a polymorphism which appears to occur throughout the species. Among Europeans the frequencies of the two alleles PGM_2^1 and PGM_2^2 are about 0.76 and 0.24 respectively, and very similar frequencies have been observed in various other ethnic groups, including various African and Oriental populations. Indeed the similar frequencies of these two common alleles in different populations suggest that the polymorphism may be remarkably stable. A number of variants due to other alleles at this locus have also been discovered, but each of these is extremely rare.

Polymorphism due to the occurrence of two common alleles at locus PGM_3 is also widespread. But here the relative incidence of the two alleles is very different in different populations. In Europeans PGM_3^1 is nearly three times as frequent as PGM_3^2, whereas among Africans the situation is effectively reversed, PGM_3^2 being about twice as frequent as PGM_3^1. Thus while the commonest phenotype among Europeans is PGM_3 1, the commonest phenotype in Africans is PGM_3 2.

At locus PGM_2 yet a further type of situation is observed. Here the great majority of individuals in all populations show the same electrophoretic

phenotype and are apparently homozygous for the common allele PGM_2^1. A number of different electrophoretic variants have indeed been demonstrated, but apart from one which is due to an allele present in many Negro populations with a frequency usually of about 0.005, all the others are very rare.

These phosphoglucomutase variants were discovered in the course of an electrophoretic screening programme deliberately aimed at searching for common polymorphic differences. The individuals studied were in the main normal and healthy, and there was no indication that the common variant types of the enzyme were associated with any marked functional differences which might be of selective significance. It seems therefore that if such differences do occur they are probably very subtle and relatively small in magnitude.

8.2. *The extent of polymorphism*

8.2.1. Enzyme surveys

The examples discussed above illustrate something of the degree of allelic variation that may occur at different gene loci. It seems that at some loci although multiple alleles may be demonstrable, there is one allele which can be regarded as the standard or normal form and is almost universally present while all the others are extremely rare. At other loci (e.g. the β-haemoglobin locus and the G-6-PD locus) although a standard allele occurs and is recognisable as such, there are in some populations but not in others alleles which are present in sufficient frequency as to give rise to a common polymorphism. At still other loci (e.g. the α-haptoglobin locus and the PGM_1 and PGM_3 loci) polymorphism is the rule. Two or more alleles occur relatively frequently and are widely distributed in many different populations. Indeed in certain cases there appears to be no valid reason for regarding one allele rather than another as the so-called normal or standard form.

A question which obviously arises is what is the relative incidence of these various situations among gene loci in general, and in particular how often do polymorphisms occur. Do the haemoglobin, G-6-PD, haptoglobin and phosphoglucomutase polymorphisms represent special and perhaps rather unusual forms of variation not typical of enzymes or proteins in general, or are they examples of a relatively common phenomenon? It would clearly be of great interest to know how many of the very large number of proteins and enzymes which are formed in the human organism exhibit this sort of variation. Since the structure of each protein is presumed to be determined by at least one gene locus, we are in effect asking at what proportion of this

vast array of gene loci do two or more relatively common alleles occur in different human populations.

In principle it should be possible to get at least an approximate answer to this question by examining in detail a series of arbitrarily chosen enzymes and proteins in randomly selected individuals and preferably in several different populations. In each case one would aim to see whether or not common polymorphism is demonstrable, and in what proportion of the enzymes or proteins it occurs.

The main results to date of a project designed along these general lines (Harris 1966, 1969) are summarised in table 8.5. A series of different enzymes chosen in an essentially arbitrary fashion were examined in population samples varying in different cases from a few hundred to several thousand individuals, using the technique of starch gel electrophoresis. The criteria used in selecting the enzymes for study were simply that they should occur in relatively accessible tissues such as blood or placenta, and that a suitably sensitive electrophoretic method could be devised by which a search for variants might be conducted on an adequate scale. There is no reason to believe that as far as the incidence of variants is concerned, the enzymes chosen were unrepresentative of enzymes in general. The individuals studied were mostly European, but samples from other population groups, mostly Africans, were also investigated though on a more limited scale. Where variant forms of an enzyme were found, detailed family studies were carried out to determine their genetical basis.

Polymorphism can be conveniently defined for the present purpose as any situation in which at a given gene locus two or more alleles occur in a population, each with frequencies greater than 0.01. Out of the 20 different enzymes examined in varying degree of detail in the course of this project, seven loci showing polymorphism of this degree were identified. The allele frequencies observed in these various cases are given in table 8.5. It will be seen that in five of the cases polymorphism was present in both the European and African populations, in one case (adenylate kinase) it was present only in the Europeans, and in one case (peptidase A) it was present only in the Africans.

From what is known about the other enzymes examined it appears unlikely that the polypeptide products of more than 27 separate loci were being scrutinised for electrophoretic differences in the study of these 20 enzymes. Thus it seems that nearly one quarter of all the loci tested in the course of these surveys of arbitrarily selected enzymes are polymorphic in each of the ethnic groups investigated. If anything, the results must underestimate the

TABLE 8.5

Results of an electrophoretic survey of twenty arbitrarily chosen enzymes in European and Negro populations. Only common alleles (frequency > 0.01) are listed. Many 'rare' alleles were also detected. (Data from M. R. C. Human Biochemical Genetics Unit, Galton Laboratory, University College London.)

Enzymes	Europeans			Negroes			References
	Allele 1	Allele 2	Allele 3	Allele 1	Allele 2	Allele 3	
Red-cell acid phosphatase	0.36	0.60	0.04	0.17	0.83	—	Hopkinson et al. (1964)
Phosphoglucomutase							
Locus PGM_1	0.77	0.23	—	0.79	0.21	—	Spencer et al. (1964a)
Locus PGM_3	0.74	0.26	—	0.37	0.63	—	Hopkinson and Harris (1968)
Adenylate kinase	0.95	0.05	—	1.00	—	—	Fildes and Harris (1966)
Peptidase A	1.00	—	—	0.90	0.10	—	Lewis and Harris (1967)
Peptidase D (prolidase)	0.99	0.01	—	0.95	0.03	0.02	Lewis and Harris (1969b)
Adenosine deaminase	0.94	0.06	—	0.97	0.03	—	Spencer et al. (1968)

Other enzymes studied were: phosphohexoseisomerase, malate dehydrogenase, isocitrate dehydrogenase, red-cell hexokinase, lactate dehydrogenase, methaemoglobin reductase, red-cell pyrophosphatase, pyruvate kinase, placental acid phosphatase, peptidases B and C, nucleoside phosphorylase, triosephosphate isomerase, and a red-cell 'oxidase'. None of these showed common electrophoretic polymorphism, though a number of rare variants were identified.

true incidence of polymorphism simply because only those involving electrophoretic differences were considered. Furthermore, the discriminative power of even this technique has been found to vary considerably from enzyme to enzyme because of technical problems, and it is quite possible that in some cases polymorphic variation has been missed.

Although the total number of different enzymes examined in this way is still not very large, the results clearly imply that quite a significant proportion of all gene loci coding for polypeptide structure are likely to be polymorphic in human populations. It is of interest that essentially similar results have been obtained in analogous enzyme surveys carried out on populations of various species of *Drosophila* (Hubby and Lewontin 1966, Lewontin and Hubby 1966, Johnson et al. 1966, Stone et al. 1968), and from the data which

are now beginning to accumulate on other species it seems that enzyme and protein polymorphism of this degree will prove to be a general phenomenon in most naturally occurring animal populations.

The ubiquity of such polymorphisms is also indicated by the variety of examples which have in recent years been discovered, apparently fortuitously, in the course of other kinds of investigation. In all, nearly thirty different human enzymes and proteins have now been shown to exhibit polymorphism in one or more of the major ethnic groups (see Appendix 2, p. 284), and it is of interest that some of these polymorphisms are not demonstrable by electrophoresis and required quite different techniques for their recognition.

In some of the enzyme polymorphisms the common electrophoretic types have been shown to be associated with significant quantitative differences in level of enzyme activity or protein concentration as determined by standard assay procedures. Particularly striking illustrations of this kind of phenomenon are seen in the polymorphisms of G-6-PD polymorphism (pp. 121–127), red cell acid phosphatase (pp. 134–140) and in serum cholinesterase (pp. 109–115). Other examples are provided by the common variants of 6-phosphogluconate dehydrogenase (Parr 1966) and of the serum protein α_1-trypsin inhibitor (Fagerhol 1969). Quite probably many of the other common structural variations in enzymes in proteins will be found to have similar effects on levels of activity or on other properties of the enzyme or protein, and such quantitative differences might be expected to be of some specific functional or metabolic significance. However, where as is no doubt often the case such effects are small in magnitude, their demonstration may be difficult and may require very specialised techniques, and such possibilities have not yet been explored in great depth in most of the polymorphisms that have so far been identified.

Of particular interest in this connexion is the possibility that in some cases the common variants of a particular enzyme or protein which make up a polymorphism may, because of their structural differences, differ in the qualititative characteristics of their functional activity, rather than in their purely quantitative effects. One probable example of this is provided by the well-known ABO blood group polymorphism (ch. 7). Here it seems tha[t] the *A* and *B* alleles result in characteristic antigenic differences, because they determine glycosyl transferase enzymes with qualitatively different specificities. The 'A' enzyme apparently specifically transfers an N-acetylgalactosamine residue from a nucleotide diphosphate compound to the end of a particular type of oligosaccharide chain, while the 'B' enzyme specifically transfers a D-galactose residue.

8.2.2. The average degree of heterozygosity

The widespread occurrence of enzyme and protein polymorphisms implies that any given individual is likely to be heterozygous at many different gene loci, and it is of some interest to enquire what the average degree of such heterozygosity may be. That is, the proportion of gene loci in a single individual at which there are likely to occur two different alleles, each specifying a structurally distinct version of a particular enzyme or protein.

A rough estimate of this parameter can be obtained from data of the kind summarised in table 8.5, which gives the estimated frequencies for alleles determining electrophoretic differences in an arbitrarily chosen series of enzymes in two very different population samples – Europeans and Negroes. For each locus at which polymorphism was detected one can determine the

TABLE 8.6

Estimates of average heterozygosity per locus from survey of 20 arbitrary chosen enzymes in Europeans and Negroes (table 8.5). Only loci where incidence of heterozygotes was > 0.01 are shown

	Incidence of heterozygotes	
	Europeans	Negroes
Red-cell acid phosphatase	0.51	0.28
Phosphoglucomutase		
locus PGM_1	0.35	0.33
locus PGM_3	0.38	0.47
Adenylate kinase	0.09	—
Peptidase A	—	0.16
Peptidase D (prolidase)	0.02	0.10
Adenosine deaminase	0.11	0.06
Average heterozygosity (detected electrophoretically) per locus, assuming 27 loci were screened.	0.054	0.052
Average heterozygosity per locus for alleles determining all structural enzyme and protein variants (i.e. electrophoretic variants $\times 3$).	0.162	0.156

proportion of the population who are heterozygous for alleles giving rise to electrophoretically different forms of the particular enzyme protein. These values are shown in table 8.6. As has been mentioned earlier, it is unlikely from what is known about the enzymes examined in this survey and from the methods used, that the polypeptide products of more than about 27 different loci in all were being scrutinised for electrophoretic differences in these investigations. If this is so, then one can obtain an estimate of average heterozygosity for alleles producing electrophoretic differences, by summing the values given in table 8.6 and dividing by 27. For the European population the average value per locus obtained in this way is 0.054, and for the Negroes it is 0.052. It is of interest that although the averages are very similar for the two population groups, the contributions from different loci that make them up differ in some respects quite widely.

If it is assumed that most of these electrophoretic variants are due to single aminoacid substitutions in the enzyme protein, then from what is known about the genetic code and about the aminoacid composition of proteins, one may expect that they probably only represent about one third of all the polymorphic variant forms that actually occur, since there is no reason to think that changes in molecular charge *per se* are likely to be specifically associated with the phenomenon of enzyme polymorphism. Thus one may reasonably infer that the average heterozygosity per locus for alleles resulting in structural differences in enzymes and proteins is about three times that calculated on the basis of purely electrophoretic differences. This particular set of data therefore suggests that the average heterozygosity per locus in Europeans is about 0.162 and in Negroes about 0.156.

The findings imply that any one individual may be heterozygous at about 16% of all loci coding for the structures of enzymes and presumably also other proteins. A very similar estimate of average heterozygosity in man was obtained by Lewontin (1967) from the available data on blood group antigens using a rather more indirect type of argument.

The total number of different gene loci concerned with defining the structures of the very many different enzymes and proteins that an individual makes is not known. It seems likely to be much more than 10,000, and may well be more than 100,000 and even the latter figure would probably only represent a few percent of the total number of genes as estimated from the DNA content of cells. However, if one takes what is perhaps the relatively conservative figure of 50,000 for the number of such 'structural' gene loci, then assuming that the average heterozygosity is indeed about 16%, one would expect that in any single individual there may be about 8,000 loci at

which there are two different alleles, each resulting in the synthesis of a structural distinct form of the particular enzyme or protein.

An important point which emerges from these general considerations is the very considerable degree of individual diversity in enzyme and protein makeup which must actually occur among the individual members of human populations. A rough idea of the extent of this can be obtained by considering together some of the enzyme and protein polymorphisms that have already been shown to exist in a single population. Relevant data on eight different enzymes which are polymorphic among Europeans are given in table 8.7.

TABLE 8.7

Enzyme individuality in the English population. Data on eight enzymes.

Enzyme	Number of alleles with frequency > 0.01	Frequency of commonest phenotype	Probability of two randomly selected individuals being of same type	References
Red-cell acid phosphatase	3	0.43	0.34	Hopkinson et al. (1964)
Phosphoglucomutase				
Locus PGM_1	2	0.58	0.47	Spencer et al. (1964a)
Locus PGM_3	2	0.54	0.45	Hopkinson and Harris (1968)
Placental alkaline phosphatase	3	0.42	0.30	Robson and Harris (1967)
Liver acetyl transferase	2	0.50	0.50	Evans and White (1964)
Adenylate kinase	2	0.90	0.82	Fildes and Harris (1966)
Serum cholinesterase				
Locus E_1	2	0.96	0.92	Kalow and Staron (1957)
Locus E_2	2	0.90	0.82	Robson and Harris (1966)
6-phosphogluconate dehydrogenase	2	0.96	0.92	Parr (1966)
Adenosine deaminase	2	0.88	0.79	Spencer et al. (1968)
Combined	—	0.018	0.005	

Since the several distinct phenotypes which make up each of these polymorphisms evidently occur independently of those of the others, it is clear that a very large number of combinations of types must occur among individual members of the population. By combining the figures given in column 3 of the table, one finds that the most commonly occurring combination of types will be found in only about 1.8 % of the population. Furthermore, from column 4 one can show that the chance that two randomly selected individuals in the population would have exactly the same combination of types is only about 1 in 200. Thus quite a high degree of individual differentiation in enzyme and protein makeup can be demonstrated with what is clearly a very limited series of examples. This must surely represent only the tip of the iceberg, and one may plausibly conclude that in the last analysis every individual will be found to have his own unique enzyme and protein constitution. Furthermore one may anticipate that this uniqueness of the individual in his basic enzyme and protein makeup is in a great variety of ways reflected in his physical, physiological and other characteristics.

8.3. Rare variants

Besides the relatively common alleles which give rise to the so-called polymorphisms, a considerable number of rare alleles determining different variant forms of particular enzymes and proteins have also been found in the course of population surveys. Thus evidence for a multiple series of rare alleles at loci determining the structures of the enzymes phosphoglucomutase (pp. 48–49), lactate dehydrogenase (pp. 43–44), placental alkaline phosphatase (pp. 30–32), peptidase A (pp. 29–30), phosphohexose isomerase (Detter et al. 1968), peptidase B (Lewis and Harris 1967), 6-phosphogluconate dehydrogenase (Parr 1966), carbonic anhydrase (Tashian et al. 1968) have been obtained in the course of electrophoretic surveys in which samples from several thousand randomly selected individuals have been examined. A similar multiplicity of rare alleles determining structural variants of a number of non-enzymic proteins such as transferrin (for review see Giblett 1969), serum albumin (Weitkamp et al. 1967) and the serum α_1-trypsin inhibitor (Fagerhol 1968) have also been observed. Numerous examples of rare variants picked up in the course of routine investigations of other enzymes and proteins could also be cited, and in fact it is beginning to appear that if virtually any enzyme or protein is examined by sufficiently discriminating methods in a few thousand individuals, one or more variants due to rare alleles at the gene locus or loci coding for the particular protein are likely to be found. The multiplicity

of rare haemoglobin and G-6-PD variants discussed earlier illustrate the same general phenomenon and indicate how many rare alleles at single loci may be demonstrable if a given protein is subjected to a particularly intensive investigation.

Thus it seems that the occurrence of a very large number of rare alleles at many different gene loci are a characteristic feature of human populations. They represent, as it were, a large pool of inherited variation, the separate components of which are extremely rare, although together they may give rise to quite a significant degree of diversity among the individual members of a population. If one allows for the fact that only a proportion of all structural variants of a particular protein will have been picked up by the screening methods (mainly electrophoretic) used in the population surveys which have so far been carried out, it seems not unlikely from the data so far collected that on average about 1 in 500 people may be heterozygous at any particular locus for one or another of the rare alleles that occur. This implies that any single individual may carry a not inconsiderable number of them (perhaps 100 or more) at various gene loci.

Since these alleles are each very rare, the variant forms of the enzyme or protein they produce have in most cases only been observed in heterozygotes. In the great majority of cases such heterozygotes appear to be quite healthy and the variant does not seem to result in any obvious deleterious effects. However, in a few cases the heterozygous state has been found to lead to a characteristic disease state. Examples of this are the rare alleles which result in structurally unstable forms of haemoglobin (pp. 23–24). Besides this, a proportion of the variants which have no obvious effects in heterozygotes, are likely because of their particular structural abnormality to give rise to some specific types of clinical disorder in the homozygous state. Thus the mutant alleles which appear to be responsible for many of the rare inherited diseases which occur, can be regarded as representing simply a part of this large class of rare alleles which are evidently ubiquitous in human populations.

8.4. Mutation, selection and drift

The incidence and distribution of the enzyme and protein variants that are observed in different human populations can be considered as representing the consequences of the operation over many previous generations of three main kinds of process. These are:

(1) Mutation, which results in the appearance of new alleles in an essentially random manner.

(2) Natural selection, which tends to eliminate those alleles which reduce the biological 'fitness' of individuals who carry them, and tends to cause the spread of alleles which increase 'fitness'. In this context 'fitness' is measured by the relative contributions which individuals of different genotypes make to the next generation. Individuals of a given genotype will, for example, be on average less 'fit' if they are more prone to die in early life so that fewer survive to become parents, or if their effective fertility is reduced for some other reason. In general, the effects of a particular allele on 'fitness' will differ according to whether it is present in the heterozygous or homozygous state, and it is important to note that alleles which are relatively uncommon will mainly occur in the population in heterozygotes.
(3) Chance effects or random genetic drift, due to the fortuitous character of the sampling process determining which of the gametes (sperm and ova) produced by members of one generation happen to give rise to the new individuals of the next.

Theory suggests that a very large number of different alleles can be generated at a single gene locus by separate mutational events. Thus at a typical gene locus with a stretch of DNA containing a sequence of, say, 900 nucleotide pairs and coding for a polypeptide chain with 300 aminoacids, the number of different alleles which can arise from separate mutations involving single base alterations alone is 2,700, since each base in the sequence may be replaced by one of three others. Because of the degeneracy of the code some 20–25% of all such mutants may not result in any change at all in the structure of the polypeptide. In perhaps 2–4% of cases, because the mutation involves the alteration of a base triplet coding for an aminoacid to one coding for chain termination, the mutant allele may be expected to result in the synthesis of a shortened polypeptide lacking a greater or less portion of the carboxyl-terminal aminoacid sequence, and in most instances this is likely to lead to a complete disruption of the protein structure. But in approximately 70–75% of cases, a mutant allele with a single base alteration will lead to the synthesis of a polypeptide differing from the original type by the substitution of one aminoacid for another at a single point in the aminoacid sequence. Here the effects on the properties of the protein will depend on the position in its three-dimensional structure of the aminoacid site which is substituted and on the particular chemical and physical characteristics of the substituted aminoacid. Other types of mutational event resulting in more drastic alterations in the DNA sequence of the gene can of course also occur. They may involve deletions, duplications or other sorts of arrangement of the base pair sequence, and although these may often no longer code for a viable protein,

in at least some cases viable proteins, though with markedly altered structures, can evidently be produced. Thus a great variety of altered forms of proteins which may differ widely in their properties and functional characteristics can result from the many different alleles which may be generated by mutations at a single locus.

In recent years there have been rapid advances in our understanding of the manner in which the structural organisation of enzymes and proteins is related to their specific properties and functional activity. This has come from the elucidation in a number of cases of the detailed three-dimensional organisation of these complex macromolecules; from comparative studies of the primary aminoacid sequences of certain proteins such as cytochrome *c* and haemoglobin in a wide variety of different species; and also from investigations of the properties of particular mutant forms in which the structural modifications have been precisely identified. One point that appears to be emerging from such studies is that a significant proportion of all the possible aminoacid substitutions which may occur in a particular enzyme or protein as a result of separate mutations may have only very minor effects or perhaps no effect at all on its functional activity. This is of obvious relevance to the interpretation in terms of population genetics of the variety of structurally distinct forms of different enzymes and proteins that evidently occur in natural populations, because it implies that the effect of some mutant alleles on the phenotypic characteristics of individuals who carry them is often likely to be very small and perhaps negligible. So that such mutants from the point of view of natural selection may be effectively neutral or near neutral (for discussion and references see King and Jukes 1969).

Of course, many other mutations, because of the particular structural alterations they produce, will result in a partial or even complete loss of functional activity of the particular enzyme or protein involved. Such mutants will, in general, be expected to be in some degree deleterious and so impose a selective disadvantage on individuals who carry them. However, it is probable that only in a minority of cases is this likely to be appreciable in the heterozygous state. Even in homozygotes its severity will vary widely according to the degree of modification in the properties, and the normal functional role of the particular enzyme or protein involved.

Presumably also there are occasional mutants which confer some degree of selective advantage at least under certain environmental circumstances. But these must be extremely unusual, if only because a random change in a complex structure like a protein is, if it results in any functional change at all, much more likely to be deleterious than advantageous.

However, quite apart from the question as to whether a particular mutant is relatively deleterious and so tends to be eliminated by natural selection, or confers some kind of selective advantage and so tends to spread, the odds against any new mutant allele persisting in a population for many generations are very considerable. The new allele will on average only be transmitted to half the children of the individual who first receives it. So there is a distinct chance that it will not be transmitted to the next generation, and the chances of its being lost are compounded in successive generations. In a reasonably large stable population where each pair of parents is on average replaced by two children who become parents in the next generation, the probability (because of chance effects alone) that a new mutant will still be present after, say, 15 generations is only about 1 in 9 (Fisher 1930). The odds in favour of the persistence of a mutant are greater if the population happens to be increasing in numbers when it appears, and are less if the population is declining. But in general the majority of new mutant alleles that occur are likely to be eliminated in the course of the next ten or twenty generations in a more or less random manner.

But because new alleles are continually being generated by fresh mutations there is always likely to be in any sizeable population a collection of rare alleles, whose composition is determined more or less haphazardly and will, in general, differ from one population to another. Wright (1966) has called the phenomenon 'polyallelic random drift'. It represents a kind of balance between the generation of new alleles by mutation, and the inevitable accidental loss of some of them and also of others already present, because of the inherent genetical structure of the population.

Now in discussing the numerous enzyme and protein variants that are found to occur in human populations, we have for convenience classified them in two main groups. One group comprises those variants which are relatively common and give rise to the so-called polymorphisms, and these appear to represent only a small fraction of all the variants that actually occur. The other group is made up of a large number of different variants which are individually extremely rare, and the allele frequencies appear in most cases to be two or more orders of magnitude less than the frequencies of the 'common' alleles. Variants due to alleles with intermediate frequencies are of course observed. But in general one has the impression that the distinction one usually makes in practice between 'common' or 'rare' variants represents a real biological dichotomy and not simply the arbitrary separation of the two ends of a continuous frequency distribution.

The general characteristics of the heterogeneous group of rare variants

which are apparently found whenever a particular enzyme or protein is examined in a sufficiently large population sample, are perhaps most easily explained in terms of the concept of 'polyallelic random drift'. That is to say, they probably represent an essentially haphazard collection of mutants which happen to have persisted in the population and whose composition is to all intents and purposes randomly determined.

If this is so, the actual number of different rare variants present in a given population and their total incidence will be mainly determined by the average mutation rate, and by what is known as the 'effective breeding size' of the population (Wright 1966, Kimura 1968). The majority of such variants are likely to be neutral or near neutral in the heterozygous state, and so natural selection will play only a small role in determining which mutant alleles happen to be present and what their incidence is. There will of course be some mutants among them which have deleterious effects in heterozygotes, and the frequencies these attain will to a greater or lesser degree be limited by selection. This effect will be particularly marked with those mutant alleles which cause a drastic reduction of 'fitness' in the heterozygous state, and such alleles will be particularly represented among those which are found to have been newly generated in immediately preceding generations. Alleles which only result in deleterious consequences in homozygotes will be very much less affected by selection, and their actual incidence and distribution will be largely the result of random mutation and drift. This perhaps explains the very considerable variations that are often observed from one population to another in the incidence of particular rare 'recessive' diseases.

8.5. Selection and the polymorphisms

Turning now to the group of alleles which give rise to the common polymorphisms, and which are in fact the main source of individual diversity in enzyme and protein structure in human populations, we find a rather different situation. At any specific locus two or more particular alleles, usually only a very limited number, are involved, and they appear to represent only a small fraction of the total number of different alleles that actually exist in the population, the others being mainly very rare.

It is widely held that differential selection, generally in the sense that heterozygotes are or were in the past at some selective advantage compared with the respective homozygotes, has been the principal factor in determining the incidence of the particular alleles that constitute the common polymorphisms. While this certainly provides a satisfying general explanation

for much of the biological diversity that we know to exist within the species and which we presume is the consequence of natural selection, it has to be admitted than in most cases there is so far not much direct evidence to support this concept. The sickle-cell polymorphism in Africa provides the best documented case, and there is suggestive geographical evidence for malaria as a selective agent in the other common haemoglobinopathies and certain G-6-PD variants. But in virtually all the many other polymorphisms which have been identified, we have very little idea what selective factors might be concerned.

The special features of the sickle-cell polymorphism are a very severe pressure of selection against the abnormal homozygote, and yet a very high frequency of the gene. This means that the selective advantage favouring the sickle-cell heterozygote must have been quite considerable in order to establish and maintain the polymorphism (see pp. 216–219). In most other polymorphisms it seems that the selective pressures if they exist must be much less marked, and it may well be that in many cases they are too small to be demonstrable with the sort of data we are currently able to collect. In fact the elucidation of the nature of the possible selective forces that may have given rise to different polymorphisms is at present among the most difficult and intractable general problems in human genetics.

One way of approaching the matter is by the direct investigation of the functional properties of the structurally distinct forms of a polymorphic enzyme or protein in the hope that this might define differences which are likely to be significant metabolically or in some other way, which could be selectively important in certain sorts of environment. In a number of cases it has been shown that common structural variants of a particular enzyme are indeed associated with marked differences in activity. But there is as yet little or no indication of the possible significance of such differences in relation to selection, except perhaps in the case of G-6-PD mentioned earlier. An obvious difficulty is that selection as it affects most individuals is presumably directed at complex physiological variables dependent on many different enzymes and proteins acting together. The important thing may therefore be the constellation of enzyme and protein phenotypes of the individual rather than the characteristics of any single one of them.

Another approach is to try and find out whether particular alleles render individuals more or less susceptible to the development of particular disorders or disabilities, especially common ones. The general method is to compare the incidence of the allele in individuals affected by the particular condition with the overall incidence in the population of which they are a

part. A now well established example of this kind of effect is the association of the ABO blood groups with certain gastro-intestinal disorders (for detailed references see McConnell 1966). Blood group A individuals are somewhat more susceptible to gastric cancer than group O individuals. Group O individuals are more susceptible to peptic ulceration than group A. However, the effects are quite small, and the natural history of the diseases in question is not such as to suggest that the particular associations so far discovered could have been the main source of the selective differential which is presumed to have established and maintained this very widespread polymorphism. Nor is it known exactly how these different blood group antigens, or the enzyme differences which apparently determine them, influence susceptibility to these particular diseases. The causal relations involved are quite obscure and presumably very indirect, and the specific associations between the polymorphic types and these particular diseases could hardly have been predicted from their known characteristics. This may well be commonly the case for other polymorphisms and other diseases, and if so the search for such associations is likely to be peculiarly difficult. There are a great variety of different common diseases, including many sorts of acute and chronic infections which would in theory have been important in the establishing of different polymorphisms, and there are an increasing number of polymorphisms which might be tested for such disease associations. But if there is no particular reason to expect one sort of association rather than another in a specific case, the discovery of significant associations is likely to be somewhat fortuitous.

Another major line of attack on the problem is primarily demographic. The aim is to categorise individuals in one or more populations in terms of the various common allelic differences that are known, and then search for differences between them in the main parameters involved in selection, such as mortality and morbidity rates at various ages, and fertility. Also by family analysis investigate possible disturbances in segregation ratios, and so on. Such data, although they may give only indirect information about specific selective factors in relation to particular polymorphisms, should in principle provide an assessment of the magnitude of any selective effects that are actually occurring in the given environmental situations. And this, of course is fundamental to the whole problem. Such surveys are, however, extremely hard to mount on a scale which is both sufficiently large as to be likely to yield significant results, and yet sufficiently detailed and exact in the determination of the various demographic parameters as to yield precise answers. So far, although much suggestive information has been obtained,

the results of even the most sophisticated surveys of this sort (e.g. Morton 1964, Neel and Salzano 1967) have mainly served to emphasise the considerable difficulties of obtaining unequivocal answers in this field.

An inherent source of uncertainty arises from the fact that the environments in which human populations live today, or even in the last few generations, are very different in important aspects from those in which they lived in the past. In particular the incidence and age distribution of mortality and morbidity and its main causes have changed, and are changing profoundly. So what may have been important selective agents in the past, and may well have shaped many of the polymorphisms that we see today, may now be of only minor or no significance. We are only looking at what is inevitably a changing situation over a very narrow period of time. Furthermore, as a general rule we have no means of knowing whether in any particular polymorphism we are dealing, as is often assumed, with a situation close to stable equilibrium due to heterozygous advantage, or with the steady increase of one particular allele at the expense of another, or with its progressive disappearance.

The main alternative to invoking selective forces as the cause of particular polymorphisms is to suppose that they have arisen fortuitously by drift. Very occasionally a particular allele which happens to be introduced into a population may spread purely by chance and come to assume quite a high frequency. This is more likely to occur if the population is small and relatively isolated. There are in fact a number of situations known where alleles determining peculiar enzyme or protein variants occur with unusually high frequencies in relatively small and isolated communities, while they appear to be absent or very rare elsewhere, and most such cases of highly localised polymorphisms are perhaps usually attributable to random drift. A probable example of this kind of effect is the variant form of serum albumin found in as many as 25% of members of a group of North American Indians known as the Naskapi (Melartin and Blumberg 1966), and in several closely related tribes. This variant does not appear to occur elsewhere in the world, although it would quite easily have been picked up by very widely used procedures. Another example is the very high incidence of a 'silent' serum cholinesterase allele in a population of Eskimos in Alaska referred to earlier (p. 116). Sometimes a small group of individuals from one population migrate elsewhere and found a new community which subsequently expands more or less in isolation. If by chance one of the founder members happens to have been a carrier of a particular rare allele then this may purely fortuitously come to have quite a considerable frequency in the new population which arises. A

striking instance of this so-called 'founder effect' is the unusually high incidence of the disease hereditary tyrosinaemia (p. 272) observed in an isolated French-Canadian population living in Quebec (Laberge 1969). The condition, which is generally fatal in early life, occurs in individuals who are homozygous for an allele which in this population may occur with a frequency as high as 0.02, but which elsewhere must be extremely rare. Pedigree studies suggest that one of the founder members of the population who immigrated from France in the 17th century was probably heterozygous for the allele.

No doubt random genetic drift or the 'founder effect' have been the principal causes of many other localised polymorphisms, and have contributed to the diversity often observed between different relatively isolated populations. But it seems less likely that drift alone can account for many of the more broadly based polymorphisms which are found in the major ethnic groups and which are spread over whole continents or subcontinents, and in some cases are evidently present throughout the species. However, chance fluctuations in gene frequencies within a population in the past, as well as the migration of groups of individuals from one place to another, could well have been responsible for many of the detailed peculiarities in the distribution and incidence of particular common alleles that are observed today.

8.6. Summarising remarks

The position that seems to have been reached can perhaps be generally summarised in the following way.

It is now clear that a very great variety of structurally distinct forms of different enzymes and proteins exist in human populations, and so extensive is the individual diversity they produce, that it seems very unlikely that any two individuals, with the exception of monozygotic twins, have exactly identical enzyme and protein constitutions.

This diversity can be attributed to the occurrence of multiple alleles at most structural gene loci, and these may be broadly speaking classified into two distinct groups. One group comprises a large number of different alleles, mostly extremely rare, whose incidence and distribution can be largely regarded as the inevitable consequence of recurrent mutation and drift. Many of these alleles probably make little difference to the viability or fitness of the individuals who carry them, but others have deleterious effects either in the heterozygous or homozygous state and give rise to a variety of rare inherited diseases and other abnormalities which we know to be a

characteristic feature of human populations. The overall incidence of such deleterious mutant alleles will of cause be limited by the effects of natural selection.

The other group of alleles, though numerically much fewer, are individually much more common. They provide the basis for the great variety of enzyme and polymorphisms which evidently occur. These are quite possibly the underlying biochemical cause of much of the inherited diversity in the physical and physiological characteristics of individuals, and also in relative susceptibilities to various diseases and other disorders. But if so we have at present, with very few exceptions, no precise idea how this is brought about in specific cases, and more particularly we know little about the exact nature of the biological causes which have led to the establishment of these common polymorphic differences.

CHAPTER 9

Gene mutations and inherited disease

A recent catalogue of inherited abnormalities (McKusick 1966) lists more than one thousand distinct clinical syndromes each of which can be plausibly attributed to the effects of a single abnormal gene. They vary greatly in manifestations and severity. Some are present at birth or appear shortly after; others may not become apparent till middle or late life. Some are inevitably progressive and fatal; others give rise to only minor disability. Any organ or tissue may be affected in some degree or another, and often quite characteristic and specific pathological changes are demonstrable. Thus this great variety of disorders encompasses virtually every branch of medicine. Judging from the rate at which new examples are currently being described in the medical literature, it seems that many more must still remain to be identified.

It is usual to classify these conditions according to whether they are inherited as so-called 'dominant' or 'recessive' characteristics, and according to whether the abnormal gene concerned is located on one or other of the twenty-two autosomal chromosomes or is sex linked, that is, located on the X or Y chromosomes. Among those so far characterised, somewhat more than half can be classified as autosomal dominant disorders and nearly forty per cent as autosomal recessive. The remainder (about 8%) are mainly X linked recessive conditions. As yet no certain example of a disease state attributable to a specific abnormal gene located on the Y chromosome has been identified.

The essential point about the so-called 'autosomal dominant' conditions is that virtually all the clinically affected individuals are heterozygous. They carry one dose of the abnormal gene which comes from one parent, and one dose of its functionally normal allele which comes from the other. Because most of the abnormal genes which give rise to such 'dominant' disorders are rare, the homozygous state has generally not been observed. It would be expected, however, that this would usually be represented by a much more

severe clinical disorder than that seen in the affected heterozygotes and quite probably often be lethal in early life.

In 'autosomal recessive' disorders, the clinically affected individuals are often homozygous and carry two doses of the abnormal gene, one derived from each of the parents. Heterozygotes with one dose of the abnormal gene and one of the functionally normal allele appear in most circumstances to be quite healthy. There may, however be two or more different abnormal genes which can occur at a particular gene locus, each producing a distinctive 'recessive' disorder in the homozygous state. Individuals heterozygous for two such alleles usually exhibit a disorder similar to what is seen in the two corresponding homozygous conditions, and if these differ in their characteristics or severity the 'double' heterozygote will generally show intermediate features. A well-known example of this is sickle-cell–haemoglobin C disease.

In the so-called 'X linked recessive' disorders, the clinical disorder occurs predominantly in males. Males having only one X chromosome will, if this carries the abnormal gene, manifest the disorder, whereas females having two X chromosomes only show the disorder if both carry the abnormal gene. If as is often the case an abnormal gene causing such a disorder is rare, the condition may never have been observed in females at all.

9.1. The molecular pathology of inherited disease

If the general theory that genes produce their effects by directing the synthesis of proteins is in its main essentials correct, then it should be possible in each of these different conditions to trace the particular constellation of clinical abnormalities that are observed, back to the effects of some specific enzyme or protein defect resulting from a single gene mutation. Indeed, a full account of the pathology of such a disease should in principle start with the details of the alteration in the base sequence of the DNA brought about by the mutation; show in what way this has modified the synthesis of the specific enzyme or protein; proceed to elucidate the secondary biochemical consequences that ensue; and finally show how these give rise to the clinical signs and symptoms that are observed.

In fact, for the great majority of inherited diseases such a complete exposition is still very far from realisation. The characteristic features that we observe in any one disorder must often represent the consequences of a very complex chain of phenomena involving interactions at many different levels of the biochemical and physiological organisation of the organism. So far it is only

in relatively few conditions that it has been possible to piece together the details of even some of the steps in the causal sequence of events.

It is not yet feasible to define the primary genetic abnormality in any particular disease by examining directly the base sequence of the abnormal gene. However, where this determines an abnormal protein which can be isolated and whose structural defect can be identified, it is often possible to deduce with some precision the likely nature of the underlying abnormality in the DNA. The haemoglobinopathies provide an extensive series of examples (see ch. 1). In sickle-cell disease, for example, the various clinical and pathological features of the disorder can be attributed to the synthesis of an abnormal haemoglobin which differs from its normal counterpart in only a single aminoacid. At the sixth position in the β polypeptide chain, valine replaces glutamic acid. Since the normal β-chain contains 146 aminoacids, and each aminoacid is coded by a sequence of three bases in the DNA, we can argue that the gene determining this polypeptide chain contains a length of DNA 438 bases long, and that the mutation giving rise to the sickle-cell allele involved the sixth triplet in the sequence, that is the 16th, 17th and 18th bases. Furthermore we can reasonably infer from what has been discovered about the genetic code and about the general nature of mutations which give rise to single aminoacid substitutions, that the specific base in the triplet which has been changed is the 17th (adenine for thymine on one of the two complementary strands of the DNA).

Similar localisations of the primary defect within the gene have been made in a number of other haemoglobin diseases. In most of these there is a single aminoacid substitution in the protein, and as with sickle-cell haemoglobin the mutation can usually be pinpointed to a specific base alteration. However, in some cases other types of protein abnormality have been demonstrated, and these imply that the mutational change in the structure of the gene must have been of a different type. For instance in the Lepore haemoglobins (pp. 80–83) the abnormal polypeptide chain has an aminoacid sequence which, in its first part, is the same as that of the δ polypeptide chain characteristically found in normal Hb A_2, but which later assumes the sequence seen in the latter part of the normal β-chain of Hb A. This is most simply accounted for by supposing that a sequence which comprises the distal part of the δ-chain gene and the proximal part of the adjacent β-chain locus has been lost from the chromosome. Since the abnormal polypeptide chain which is now defined by the new gene contains 146 aminoacids, as do both the normal β-chain and the normal δ-chain, we may infer that the primary genetic abnormality is a deletion of a chromosomal segment involving a stretch of DNA at least 438

bases long. Other abnormal haemoglobins characterised by the loss of a portion of the polypeptide sequence are Hb Freiburg and Hb Gun Hill (pp. 87–88). In the former a single aminoacid is missing from the β-chain and in the latter a sequence of five aminoacids is missing indicating a deletion of fifteen bases.

Clearly different kinds of mutational events can result in the abnormal genes which cause disease. Little is known about the nature of the phenomenon (or phenomena) by which single base alterations come about. But deletions can probably arise in at least two different ways. One involves unequal crossing over between homologous chromosomes following mispairing at meiosis. This is almost certainly the cause of the Lepore deletions, and may well also account for the other haemoglobin deletions that have been identified. The phenomenon is particularly likely to occur where, as a result of some previous localised duplication of the genetic material, very similar DNA sequences lie close together on the same chromosome. The more extensive the regions of homology, the greater is the probability of mispairing, so that presumably some genes are likely to be much more prone to this type of mutational event than others. Another way in which deletions originate is when by chance two chromosomal (or chromatid) breaks occur more or less simultaneously and this is followed by aberrant reunions, so that an intervening segment is lost if the breaks occur on the same chromosome, or there is a translocation of material from one chromosome to another associated with some loss if different ones are involved.

The data obtained from structural studies on abnormal haemoglobins make it clear that both mutations involving single base changes and mutations involving deletion can lead to clinical abnormality, and presumably the same is true for other proteins and enzymes. But it is not possible as yet to assess what might be the relative importance of these or other types of mutational change as causes of the great variety of inherited diseases that are observed. Present information is derived entirely from situations where it has been possible to isolate and characterise structurally the abnormal protein, and these represent a very limited and probably somewhat biased sample of inherited disorders taken as a whole. Even in some of the haemoglobin diseases, for instance the thalassaemias (pp. 97–106) it does not appear that studies of the protein structure *per se* can provide specific information about the nature of the mutational change in the DNA. In these cases it seems that investigations in depth of the abnormality in the mechanism of protein synthesis, as well probably as structural studies at the RNA level will be necessary before insight into the type of mutation involved will be obtained.

In conditions where a specific abnormal protein or enzyme can be identified, the characterisation of its physico-chemical properties is an important step in elucidating the pathological process. In the case of sickle-cell haemoglobin, for instance, the critical phenomenon is its dramatically reduced solubility in the deoxygenated state. This is presumably a consequence of altering a very small region of the surface of the protein molecule by substitution of the hydrophobic valine for a hydrophilic glutamic acid residue (p. 18). It is a striking fact that although many other types of aminoacid substitution occurring in different parts of the haemoglobin molecule have been identified, the very marked change in solubility occurring in sickle-cell haemoglobin still remains unique to this substitution. The altered solubility is the reason for the morphological changes seen in the red cells when they are exposed to low oxygen pressure, the so-called sickling phenomenon. *In vivo* similar deformation of the red cells tends to occur in the venous side of the circulation particularly in the small veins and venous capillaries, leading to increased blood viscosity. In turn these effects can result in localised thrombosis and tissue damage. Also the deformed red cells are more readily destroyed than normal ones, and so chronic anaemia occurs and this induces other secondary pathological changes. Thus, although many of the details are still obscure, it is possible to envisage the main sequence of events which give rise to the clinical syndrome of sickle-cell anaemia. The initial mutational alteration involves only a very small change in the DNA of the gene, but its effects are progressively amplified first by the subtle alteration in haemoglobin structure which modifies its solubility properties, and then by the effects this has on the characteristics of the circulating red cells, so that eventually a complex pattern of pathological changes is produced.

One of the physical properties of a protein which may be significantly altered by a slight modification in structure is its stability. If an abnormal protein is markedly less stable than its normal counterpart, then the rate at which it is denatured *in vivo* is likely to be much increased, and the loss of functional activity which results can have important pathological consequences. For example, several inherited forms of severe chronic anaemia have been shown to be associated with abnormal haemoglobins whose most striking feature is their instability (pp. 23–24). They undergo much more rapid denaturation in the red cell than normal haemoglobin and this is evidently the main cause of the various pathological effects that are observed. Several abnormal enzyme proteins, for instance the glucose-6-phosphate dehydrogenase variant Gd Mediterranean (pp. 126–130), also exhibit decreased

stability and here again such secondary biochemical and clinical disturbances that ensue can be largely attributed to this effect. Probably many other inherited disorders have a similar causal basis.

Instability is likely to be brought about by any change in the primary structure of a protein which significantly distorts its normal three dimensional conformation. Where this is due to a single aminoacid substitution the severity of the effect will depend on the chemical properties and size of the side chain of the aminoacid which is substituted, and also the precise site of the substitution. However, a number of quite different substitutions occurring at different sites in the protein molecule may well have essentially the same consequences as far as their effect on the stability of the protein is concerned, and thus give rise to the same pathological process. Consequently a variety of different mutations may cause a series of distinct conditions which are, however, in all respects other than the primary structure of the abnormal protein, indistinguishable from one another. It may also be noted that different small deletions within a gene may, by leading to the absence of one or several aminoacids and consequent shortening of the corresponding polypeptide chain, each result in marked distortion of three-dimensional structure with much reduced stability of the protein. Larger deletions of course, by causing even greater abbreviation of the polypeptide, would often fail to lead to the appearance of a recognisable protein at all. Similar effects could also result from those mutations causing single base alterations where the base change alters a base triplet coding for an aminoacid to one coding for chain termination.

In the study of the properties of abnormal enzyme proteins, investigation of the kinetics of their catalytic activity is a matter of obvious interest. Changes for example in affinity of an enzyme protein for substrate or coenzyme which would be reflected by altered kinetics, may clearly be important causal factors in the development of a clinical disorder. Examples are the altered kinetics of argininosuccinate synthetase in citrullinaemia (pp. 159–160), of the 'atypical' form of serum cholinesterase associated with suxamethonium sensitivity (pp. 110–115), and of the glucose-6-phosphate dehydrogenase variant Gd Oklahoma (pp. 130–131) which causes a particular form of chronic haemolytic anaemia. In each of these cases the Michaelis constants (K_m) with respect to the enzyme substrate have been shown to be significantly elevated and the magnitude of the effect seems to be sufficient to account for the pathological consequences observed in these conditions.

One would expect that quite a number of different single aminoacid substitutions in an enzyme protein would either by causing an alteration in the

conformation or the chemical structure of the active site, result in one way or another in a change in the kinetic parameters. The same substitution might also lead to an alteration in one or more of the other physico-chemical properties of the enzyme protein, for example its molecular stability. It is of obvious importance, therefore, in the elucidation of the pathology of a particular condition to assess the relative significance of such different effects. An interesting illustration is provided by the glucose-6-phosphate dehydrogenase variant, Gd Mediterranean. Although the details of the structural alteration in the abnormal enzyme protein are not yet known, several significant changes in its properties have been recognised. It is much less stable than its normal counterpart, the Michaelis constants, with respect to both the substrate glucose-6-phosphate and the coenzyme NADP are lower than normal, and it shows an increased facility to utilise the substrate analogue, 2-deoxyglucose-6-phosphate. The clinical disorder, favism, with which this abnormal enzyme is associated, almost certainly occurs because of the very low level of the enzyme activity present in the abnormal red cells, and this can be largely accounted for by the marked instability of the enzyme protein. The altered kinetics are probably of only minor or no significance in the development of the pathological process. Indeed, reduced Michaelis constants for substrate and coenzyme would be expected to be associated with enhanced activity.

Many different mutations may lead to the specific deficiency of a particular enzyme, either by causing the synthesis of an abnormal enzyme protein with altered kinetics or stability, or by causing a true reduction in the rate of synthesis of the enzyme protein, or a complete failure in synthesis; and in quite a number of inherited diseases it has been possible to identify such a specific enzyme deficiency as central to the pathology of the condition, even though the precise molecular basis of the enzyme abnormality is not yet understood. Such conditions are for historical reasons usually referred to as 'inborn errors of metabolism', though in principle this name might well be applied to virtually any inherited disease. Various examples have already been discussed (p. 141) and others are listed in Appendix 1 (p. 266).

Typically an unusual but quite distinctive pattern of biochemical changes is observed, characterised by abnormally increased concentrations of certain substances in the body fluids or intracellularly, and by the relative deficiency of others. The detailed character of these changes will in general depend on the role of the particular enzyme in normal metabolism and on its tissue localisation. Their magnitude and, to some extent, their distribution will depend on the degree to which the specific enzyme activity has been reduced.

In some conditions it may be effectively absent, but in others the reduction in activity is less extreme.

The clinical abnormalities seen in these various metabolic disorders presumably represent secondary consequences of the distorted biochemical pattern set up by the specific enzyme defect. But the detailed causal relationships are often difficult to discern. Even in such an extensively investigated condition as phenylketonuria (pp. 145–148), where a great deal is now known about the character of the metabolic disturbance, and much information about the altered concentrations of a wide variety of metabolites has been obtained, it has still not proved possible to obtain a satisfactory explanation for the severe mental retardation which is the outstanding clinical feature. The neurones of the developing brain are evidently affected in some way by the biochemical upset, but the details of the process are still quite obscure.

It is perhaps worth commenting on the unexpected character of the biochemical disturbances which have often been discovered to underlie particular clinical syndromes. The point is well illustrated by the condition known as homocystinuria (p. 273). The clinical syndrome is complex and includes among its characteristic features such diverse abnormalities as mental retardation, dislocated lens, a tendency both to arterial and venous thrombosis, and abnormalities in the development of the bones. An abnormal urinary excretion of homocystine was discovered to be associated with this syndrome in the course of a routine screening programme of the aminoacids in the urine of mentally retarded patients. This led to the recognition of a disturbance in methionine metabolism due to a specific deficiency of the enzyme cystathionine synthetase (L-serine dehydratase). Although a great deal was known about the metabolic pathways involved in the conversion of methionine to cysteine prior to the discovery of homocystinuria, biochemists would hardly have predicted that a block in this pathway might give rise to the particular complex of clinical abnormalities that are observed. Indeed it is still not known how the various clinical features are brought about, though there is little doubt that they all in some way stem from the primary deficiency of cystathionine synthetase.

The condition known as Pompe's disease (p. 166) illustrates the same general point. This had long been recognised as a quite characteristic disorder in which there was progressive accumulation of glycogen particularly in heart muscle. But although the main pathways in the synthesis and degradation of glycogen had apparently been established, and the various enzymes thought to be concerned had been examined in Pompe's disease, the underlying nature of the condition had remained obscure. Indeed the discovery

that it is due to a specific deficiency of an α-(1,4) glucosidase, normally present with many other hydrolases in the intracellular organelles known as lysosomes, was quite unanticipated because this enzyme was not previously thought to have any significant role in glycogen degradation.

When one considers such examples it is perhaps not surprising that there remain very many inherited disorders in which we still have virtually no idea what enzyme or protein may be defective, or indeed what area of metabolism could be involved. The clues may well be present in the symptomatology, but they are certainly in most cases far from obvious and as yet no one appears to have thought of the right biochemical systems to investigate.

9.2. Dominance and recessivity

With the progressive elucidation of the molecular pathology of a disorder one may expect to see more clearly the reasons for its particular mode of inheritance; that is why it should occur predominantly in the heterozygous state and so be classified as a 'dominant', or why it should be seen only in the homozygous state and so be classed as a 'recessive'.

Sickle-cell disease which is inherited as a 'recessive' characteristic provides a simple illustration of the point. The clinical abnormalities in this condition derive from the fact that the haemoglobin present in the red cells is extremely insoluble when deoxygenated and so causes red cell deformation (sickling) *in vivo*, in those parts of the circulation where the oxygen tension is low. In heterozygotes for the sickle-cell gene, about 65% of the haemoglobin present in the red cells is usually of the normal type and only about 35% is of the sickle-cell type, though the total amount of haemoglobin per cell is not significantly reduced. The mixture considered as a whole shows a reduction in solubility when compared with normal haemoglobin, and sickling of the red cells can be readily demonstrated *in vitro* if the oxygen tension is sufficiently reduced. But the degree of deoxygenation required is greater than normally occurs *in vivo*. So untoward consequences do not usually occur in the heterozygotes who are generally quite healthy.

This example emphasises the important point that the terms 'dominant' and 'recessive' as generally used, only have meaning in reference to a specific characteristic or phenotype. Sickle-cell disease is inherited as a 'recessive' characteristic; but the sickle-cell phenomenon, that is the occurrence of sickling of the red cells when they are subjected to appropriate *in vitro* procedures, is inherited as a 'dominant' characteristic since it occurs both in heterozygotes and homozygotes for the abnormal gene.

One may usefully contrast 'recessively' inherited sickle-cell disease with the 'dominantly' inherited forms of chronic anaemia due to the so-called unstable haemoglobins. In these conditions the heterozygote synthesises in his red cells both the unstable haemoglobin and also the normal haemoglobin, but because of its rapid denaturation the amount of the unstable form and hence the total amount of haemoglobin present in a functionally active state, is progressively reduced as the red cells mature. This causes a chronic anaemic state which is further accentuated by precipitation of the denatured abnormal haemoglobin, which tends to cause the red cells to be more rapidly destroyed and removed from the circulation. Thus unlike the situation in the sickle-cell heterozygote the normal haemoglobin synthesised by these heterozygotes fails to protect the red cells from the adverse effects of the abnormal form. The genes determining the various types of unstable haemoglobin are all extremely rare and so the disorders produced in the homozygous states have not been observed. One may predict however that they would be extremely severe forms of anaemia, often lethal in early life.

In diseases arising from the deficiency of a specific enzyme, the question as to whether the disorder is inherited as a 'recessive' or as a 'dominant' may largely depend on the average level of activity of the enzyme in the normal homozygous state, and in particular on how far this is in excess of the minimal level actually needed to maintain healthy function. In the heterozygous state the level of the specific enzyme activity is generally intermediate between that present in the normal and in the abnormal homozygotes, and in the extreme case where the abnormal gene leads to complete loss of the enzyme activity, the heterozygote usually shows about one half the average activity found in normal homozygotes (pp. 171–177). So if, as usually appears to be the case, the level of activity in the normal homozygotes is on average many times the minimum actually required for efficient metabolic function, the reduced activity present in heterozygotes will also usually be in excess, and no untoward consequences will result. Clinical abnormality will then only be seen in the abnormal homozygotes where the activity is presumably so diminished that it is insufficient to maintain healthy function. In fact most of the so-called 'inborn errors of metabolism' which have so far been identified have a recessive pattern of inheritance and one may infer that the normal levels of the specific enzymes involved are much higher than is strictly required. There is, as it were, a considerable built in safety factor.

Dominant inheritance of a disease due to an enzyme deficiency is most likely to occur where the enzyme in question happens to be rate limiting in the metabolic pathway in which it takes part, because the level of activity of

such enzymes in the normal organism will in general be closer to the minimum required to maintain normal function.

9.3. Heterogeneity of inherited disease

Very often a particular syndrome which at first was thought to be a discrete entity and presumed to be determined by a single abnormal gene has turned out on closer analysis to represent a collection of distinct disorders, each the consequence of a quite different mutation and each with its own specific underlying pathology. The degree of genetical heterogeneity that may be uncovered in what at first sight seems a relatively homogeneous condition may be quite remarkable, and there seems no doubt that this will prove to be a very widespread and general phenomenon.

That several quite different abnormal genes may often result in clinical disorders which closely resemble one another or indeed appear identical is not surprising. Loss of function of one or another of a series of enzymes which are involved in a consecutive sequence of reactions in a metabolic pathway, or which are associated together in a complex of physiological relationships, could well result in the same or very similar end results at the clinical level. So a number of quite distinct abnormal genes, each affecting a different enzyme, may all give rise to closely similar clinical consequences. Furthermore, defective function of any given enzyme or protein may arise because of structural alterations in distinct polypeptide chains coded by separate gene loci, or may be the consequence of a mutation at some other locus specifically concerned with regulating the rate of synthesis of the protein. Finally, at any one locus mutational changes occurring at quite different sites of the DNA sequence, though affecting the structure of the corresponding polypeptide chain in different ways, may each result in loss of function and so give rise to the same or very similar clinical manifestations. Thus what appears to be a single disease entity may in fact be brought about by genes at several quite different loci, and also by different alleles at a single locus.

Congenital methaemoglobinaemia (pp. 19–23) provides a simple and well documented example. The characteristic clinical feature of this rare syndrome is a greyish blue cyanotic appearance which is due to the fact that a significant fraction of the iron in the haemoglobin in the circulating red cells is in the ferric state and incapable of transporting oxygen. The abnormality is apparent at birth or shortly after, and persists throughout life usually with very little variation. Most affected individuals are not seriously incapacitated, though occasionally the abnormality is associated with some degree of

mental retardation. The condition needs to be distinguished from other causes of chronic cyanosis, such as congenital malformations of the heart, but this can usually be done relatively easily on clinical grounds.

The early studies on the inheritance of the condition showed that at least two sorts of genetical abnormality occurred (Barcroft et al. 1945, Hörlein and Weber 1948). In some cases the disorder appeared to be inherited as an autosomal recessive condition, while in others it was apparently inherited as an autosomal dominant. In the autosomal recessive cases a specific deficiency of the red cell enzyme methaemoglobin reductase was identified (Gibson 1948, Scott and Griffith 1959). In affected individuals, the level of this enzyme activity is consistently very low and often only barely detectable. Among their parents, children and certain other relatives, who may be presumed to be heterozygotes, a partial reduction in the level of activity of this enzyme is found (Scott 1960), but this is evidently insufficient to lead to any significant degree of methaemoglobinaemia or cyanosis. Electrophoretic studies on the residual enzyme activity in certain affected individuals suggest that, at least in some cases, a structural defect in the enzyme protein is probably the cause of the enzyme deficit, but that the structural alteration differs in patients from different families (Kaplan and Beutler 1967, West et al. 1967). So probably several different abnormal alleles at the gene locus (or loci) coding for the methaemoglobin reductase enzyme protein can bring about this condition.

In the group of 'dominantly' inherited forms of congenital methaemoglobinaemia, methaemoglobin reductase is not abnormal. However, in most of these cases there is a structural defect of the haemoglobin itself, and a number of distinct types of abnormality have been recognised. They each involve a specific single aminoacid substitution occurring in a region of the molecule where a haem group is attached to a polypeptide chain, and either the α- or the β-chains which are of course coded at separate gene loci may be affected. A minor difference in clinical manifestation is that in the α-chain mutants the cyanotic appearance is present at birth because the α-chain occurs in both foetal ($\alpha_2\gamma_2$) and adult ($\alpha_2\beta_2$) haemoglobin, whereas in the β-chain mutants it only becomes apparent some weeks after birth when adult haemoglobin begins to be the predominant form present.

Thus the syndrome of congenital methaemoglobinaemia can arise because of mutations at at least three distinct gene loci. One is concerned with defining the structure of methaemoglobin reductase, and the others with defining the α- and β-chains of haemoglobin, and at each of these loci several different abnormal alleles causing the syndrome evidently occur. That mutations at

yet other gene loci may also result in the same sort of clinical disorder is indicated by the report of a form of congenital methaemoglobinaemia apparently inherited as an autosomal dominant, but showing no abnormality in either haemoglobin or methaemoglobin reductase (Townes and Morrison, 1962).

Congenital methaemoglobinaemia is an extremely rare condition, and probably in most populations occurs with an incidence of only perhaps one in several hundred thousand births. Nevertheless at least eight different abnormal genes causing it have already been identified, and the syndromes they separately produce are difficult if not impossible to distinguish on clinical grounds alone. One may anticipate that a similar degree of genetic heterogeneity is likely to be found in many other inherited conditions which at present can only be defined in clinical terms, or in terms of some of the secondary biochemical or physiological disturbances originating from the underlying enzyme or protein abnormality.

Phenylketonuria, for example, is a disorder in which the various characteristic biochemical changes observed in the blood, cerebrospinal fluid and urine can be quite clearly attributed to a failure of the normal metabolic conversion of phenylalanine to tyrosine due to a deficiency of the enzyme phenylalanine-4-hydroxylase (pp. 145–148). The condition is inherited as an autosomal 'recessive' disorder, and it has usually been assumed that the affected patients are all homozygous for the same abnormal gene. But clearly a number of different abnormal genes, which result in different ways in the loss of the specific enzyme activity could well occur. If so, the affected patients might include individuals homozygous for one or other of these genes, and also individuals heterozygous for two of them. One may also note that mutations at two or more different gene loci might be causes of the condition. At present there is insufficient information about the structure of the enzyme protein or the regulation of its rate of synthesis to allow us to assess the likelihood of this possibility. Nor have any linkage markers been discovered which could be useful in indicating whether genes at different loci do in fact cause the disorder.

The various mutations which give rise to the deficiency of a specific enzyme may, of course, vary from one to another in the degree of enzyme deficit they produce, and hence in the severity of the clinical abnormality that occurs. In some cases no functionally active enzyme may be formed at all, so that the different mutants from which this results may all cause what amounts to the 'same' disease. But where enzyme activity is not completely abolished, the degree of deficit produced by different mutants can vary

considerably, and so the 'inherited disease' apparently caused by a particular abnormal gene may appear to vary very widely in its manifestations and severity, until the underlying molecular and genetical differences between different cases have been defined. Much of the commonly observed clinical variation in what are considered to be single disease entities almost certainly stems from this general phenomenon.

The wide variation in functional activity and the range of clinical effects that may be produced by different mutants affecting a given enzyme protein is particularly well illustrated by the numerous variants of G-6-PD. Nearly thirty distinct variant forms of this enzyme protein have been identified (pp. 121–131). They have been distinguished from each other and from the normal enzyme by such criteria as electrophoretic mobility, Michaelis constants, thermostability and *p*H optima, and they appear to be determined by a series of alleles each of which produces a distinctive structural change. Since the alleles occur at a gene locus on the X chromosome it has been possible in male subjects to examine the effects of each of these alleles on the level of activity of the enzyme separately. Most of the studies have been on red cells, and here it has been found that the average G-6-PD activity in cells containing the different variants varies considerably from one to another. In some only traces of activity are detected. In others a less severe but nevertheless quite marked reduction in enzyme activity is present. There are also others in which the average level compared with the normal is only slightly reduced or not at all. These differences in activity level are in many of the cases attributable to differences in stability of the enzyme proteins. But in some cases alterations in the kinetic or other characteristics may be important.

In several of these variants the degree of enzyme deficiency and the consequent disturbance in red cell metabolism is sufficiently severe as to result in a state of chronic haemolytic anaemia because of premature destruction of the red cells. In other cases affected individuals are healthy under most circumstances, but are particularly liable to develop haemolytic attacks if exposed to what are for them, noxious agents. Thus individuals with the variant Gd Mediterranean are sensitive to some substance present in fava beans, and the variant Gd A− causes the well-known sensitivity to primaquine and certain sulfonamides which is common among Negroes. There are still other G-6-PD variants which are apparently quite harmless, and in general the wide range in the degree of enzyme deficit associated with these different variants is roughly correlated with the severity of their clinical consequences.

Similar relationships have been shown with defects of other enzymes. A particularly interesting example concerns the enzyme hypoxanthine-guanine

phosphoribosyl transferase, which in the normal individual is involved in the regulation of uric acid formation (fig. A 7, p. 282). Lesch and Nyhan (1964) described a complex and very severe neurological syndrome in children which is characterised by mental retardation, spastic cerebral palsy, choreoathetosis, and a curious behavioural disorder manifested by self-destructive biting. This was found to be associated with hyperuricaemia due to overproduction of uric acid, and in some cases uric acid renal calculi and signs of gout are present. The condition turned out to arise from what appeared to be an apparently complete and quite specific deficiency of the enzyme hypoxanthine-guanine phosphoribosyl transferase (Seegmiller et al. 1967), determined by an abnormal gene located on the X chromosome (Nyhan et al. 1967).

Because of its relationship to uric acid formation, the enzyme was then examined in a number of adult patients with typical clinical histories of acute gouty arthritis or uric acid nephrolithiasis and with hyperuricaemia shown to be due to uric acid overproduction (Kelley et al. 1967, Kelley 1968). Among this series of patients some were found to have quite marked deficiencies of the enzyme, though not as severe as in the Lesch–Nyhan syndrome. The results of some of the enzyme assays are shown in table 9.1. They suggest that several quite distinct types of defect may occur and give rise to hyperuricaemia, but in any one family the same specific abnormality is present among the affected individuals. Thus in one family (table 9.1, family J) the affected individuals showed levels of activity as measured in red cells of only about 1% of the normal when either hypoxanthine or guanine was used as substrate. Furthermore, the enzyme protein was found to be significantly more thermolabile than the normal enzyme. In another family (table 9.1, family L) the affected individuals also showed much reduced enzyme activity, but the reduction was considerably more marked with guanine as substrate than with hypoxanthine, a finding which indicates an altered pattern of substrate specificity. Furthermore here the enzyme protein appeared to be less thermolabile than the normal one.

These findings suggest that several distinct abnormal genes, each producing a structurally altered form of this enzyme protein with abnormal properties, separately occur in the various families. They each evidently lead to marked enzyme deficiency resulting in overproduction of uric acid with hyperuricaemia, and the clinical consequences (gout and nephrolithiasis) are very similar in the different cases. But the clinical picture contrasts very strikingly with that of the Lesch–Nyhan syndrome which is apparently the consequence of a

TABLE 9.1

Hypoxanthine-guanine phosphoribosyl transferase activity assayed with hypoxanthine or guanine as substrate in red cells. (Kelley 1968).

	Phosphoribosyl transferase activity (mμmoles/mg protein/hr)		Heat stability relative to normal
	Hypoxanthine	Guanine	
Control subjects (18)	103±18	103±21	—
Patients with Lesch-Nyhan syndrome (9)	<0.01	<0.004	—
Patients with gout due to uric acid overproduction			
(a) with normal enzyme (10)	99±13	106±21	—
(b) with defective enzyme (10)			
J family			
FJ	1.3	0.6	reduced
RJ	1.5	0.8	reduced
TJ	1.8	0.8	reduced
L family			
FL	11.8	0.5	increased
ML	8.7	0.5	increased
S_1 family			
TS	9.9	9.5	not tested
D family			
AD	12.2	17.3	not tested
G family			
JG	9.4	8.8	not tested
RG	9.2	7.5	not tested
S_2 family			
GS	0.03	0.009	not tested

complete or almost complete deficiency of the enzyme. Here a very severe neurological disorder manifesting in childhood occurs.

In conclusion it should also be noted that variation in the expression of a particular inherited disease in different individuals may be derived not only from differences in the particular mutant genes causing the specific enzyme or protein defect which is the basic cause of the disease in question, but also from differences in the rest of the genetical constitution of the individuals in whom the particular abnormal gene occurs. Probably no two individuals,

with the exception of monozygotic twins, are exactly alike in this respect, because of the multiplicity of different 'normal' allelic combinations that are present at many gene loci (pp. 225–232). Such differences in so-called 'genetical background', because they define many of the details of the biochemical and physiological milieu against which the effects of the particular abnormal gene are expressed, may well influence the manifestations of a particular disorder. Some combinations of genes at other loci might minimise the pathological consequences, and others accentuate them. Since such different combinations of genes may often result in little or no obvious differences between individuals not carrying the particular abnormal gene, the detailed manner in which they contribute to the variation is generally difficult to define.

In this connexion it is necessary to distinguish between variability due to alleles at loci other than the one where the abnormal gene whose effects are being considered occurs, and variability due to the existence of different so-called 'normal' alleles at the same locus. The latter source of variation can of course only exist where the affected individuals are heterozygous (i.e. in 'dominant' diseases). It has been shown that several distinct alleles having different effects on function without necessarily resulting in overt disease may occur at a particular gene locus, and that sometimes two or more such alleles are each relatively common. In these circumstances the abnormal gene, whose effects one is considering, may be present in heterozygous combination with one or another of these different 'normal' alleles, and this could influence the expression of the clinical disorder that ensues. This effect is sometimes termed 'allelic modification'. It tends to result in affected sibs resembling one another more closely in the manifestations of the particular disease than affected parents and children.

9.4. Heredity and environment

The characteristics of an individual depend not only on the genes he carries and the types of enzymes and proteins he consequently makes, but also on the environment or more correctly the series of environments in which he develops and in which he lives. The interactions between what Francis Galton called 'nature' and 'nurture' are often complex and difficult to disentangle. However we implicitly recognise their importance in our ideas of what we mean by inherited disease, and the question can perhaps be examined most simply by looking at a few selected diseases in which the essential nature of the genetical abnormality at the enzyme level has been

TABLE 9.2

Genetically determined enzyme deficiencies and environmental factors, in the causation of different disorders.

Condition	Enzyme deficiency	Environmental factor
Galactosaemia	Galactose-1-phosphate, uridyl transferase	Lactose, galactose
Hereditary fructose intolerance	Liver aldolase	Fructose, sucrose
Favism	Glucose-6-phosphate dehydrogenase	Fava bean
Primaquine sensitivity		Primaquine, sulfonamides, etc.
Suxamethonium apnoea	Serum cholinesterase	Suxamethonium
Scurvy	L-gulonolactone oxidase	Vitamin C deficiency

defined, and also the principle environmental factors affecting its expression are more or less understood (table 9.2).

Consider galactosaemia (pp. 148–149). In this condition there is a specific inability to metabolise the sugar galactose because of a genetically determined deficiency of the enzyme galactose-1-phosphate uridyl transferase. Galactose in the form of the disaccharide lactose is the major carbohydrate constituent of milk, and so in the usual course of events newborn infants will receive large amounts of it. If the transferase enzyme is lacking, there is a block in metabolism. Galactose accumulates in the blood, galactose-1-phosphate accumulates intracellularly, and the clinical consequences are severe. The infant fails to thrive, weight gain is slow, and eventually liver and brain damage and cataract formation occur. If however a diet free of galactose but adequate in other respects is fed to the infant, a dramatic improvement occurs. Indeed it appears that if the diagnosis is made sufficiently early before irreversible damage has occurred, the infant will develop in a normal and healthy manner.

In a sense then, galactosaemia can be regarded as an inborn inability to cope with one particular facet of an infant's normal environment, namely lactose in milk. If the environment is modified appropriately the ill effects of

the disability can be prevented. The individual will still lack the particular enzyme, but in his new environment this will not incommode him. He will as it were be predisposed without being clinically affected.

The condition known as hereditary fructose intolerance (pp. 151–154) illustrates the same general point. Here there is a deficiency of the enzyme liver aldolase (aldolase B) which is necessary for the normal metabolism of fructose. If fructose or sucrose is fed, rather severe symptoms follow. If fructose is excluded from the diet no harmful effects occur. Unlike galactose, fructose is not an obligatory feature of the normal diet of an infant. Consequently the manifestation and severity of the disease are more variable. An important factor appears to be the time when breast feeding is discontinued. If an early change to artificial milk feeds with added sucrose is made, rapid deterioration in the infant's condition occurs and should the nature of the abnormality not be recognised, irreversible damage may take place. If, however, the infant is not weaned until he is several months old he is often able to make a positive rejection of feeds which make him ill, and the peculiarity is likely to be diagnosed. Furthermore such children tend to develop a strong aversion to sugar, sweets and fruit, and thus tend to protect themselves. This illustrates rather clearly how the severity of an inherited disease may depend on quite fortuitous and apparently unconnected factors in the infant's environment. In this case the time when breast feeding is discontinued. In passing one may also note that it also illustrates, in a somewhat dramatic manner, the way in which an individual's taste and food preferences may be a direct consequence of a specific enzyme abnormality.

A particularly clear example of the way environmental and genetical factors may interact in the causation of a disease process is provided by favism. This has been known for many years as a severe form of recurrent haemolytic anaemia in which the attacks follow the ingestion of the fava bean. It is found particularly in the Middle East and certain parts of Southern Europe where the fava bean is a common feature of the diet. However by no means all people who eat fava beans develop favism. Those that do have been shown to have a genetically determined deficiency of G-6-PD usually due to the variant known as Gd Mediterranean. So to develop the disease an individual must carry the gene and eat the bean. Both genetical and environmental factors can thus be clearly seen as essential to the causation of the disease. In fact the matter is even more complex because not all individuals with the specific form of enzyme deficiency who eat fava beans develop overt haematological disease, and among those that do its severity can be variable. The detailed reasons for this are not known, but there is some evidence that other

genetical factors may be involved as well as differences in the quantity and the form in which the beans are eaten.

We have little hesitation in regarding galactosaemia as a hereditary disease because it is manifest in all individuals with the appropriate genetical constitution. The environmental feature to which they are ill adapted is universally present. The same is essentially true for hereditary fructose intolerance except that here the environmental factor is somewhat more variable and this is reflected by variation in the severity of the clinical condition. It is not however possible to regard favism as purely genetical or purely environmental in origin, and this is probably true of many other conditions that occur in clinical practice.

In fact whether a disease tends to be regarded as genetical or environmental in origin largely depends on the relative prevalence of the genetically determined predisposition to the condition on the one hand, and of the particular environmental situation which elicits it on the other. If the genetical predisposition is relatively rare, and the significant environmental factor is common or indeed universal – as is the case with galactosaemia – we say that the disease is inherited. If however the genetical predisposition is relatively common and the unfavourable environmental situation occurs infrequently, then the environmental factors appear as the most important causal agents.

An extreme example is scurvy. As far as we know all members of our own species and also other primates are incapable of synthesising L-ascorbic acid (vitamin C). Consequently if for some reason their diet is deficient in this substance they develop scurvy. This is not the case in many other mammalian species which can apparently manage without an independent dietary source of L-ascorbic acid, because they can synthesise it from D-glucose via the reaction sequence:

D-glucose → D-glucuronolactone → L-gulonolactone → L-ascorbic acid.

Evidently in man and other primates the enzyme activity capable of converting L-gulonolactone to L-ascorbic acid is lacking (Burns 1957) presumably because the necessary gene has been lost in the course of evolution. So scurvy with good reason is considered to be a disease caused by an unfavourable environment, namely one in which the individual's diet is deficient in vitamin C. But it can equally well be thought of as due to an inborn error of metabolism which we all happen to possess (Snyder 1959).

In some cases of course, both the inherited predisposition to a particular disease, and also the specific environmental factors which elicit it, may both

be relatively uncommon. In these circumstances the disease is likely to be rare and very irregularly distributed, and it may not be at all apparent from family studies that hereditary factors are involved at all. A simple model of this kind of situation is the occurrence of the abnormality known as 'suxamethonium apnoea' (pp. 109–119). About 1 in 2,000 people are excessively sensitive to the drug suxamethonium which is often used to obtain muscular relaxation during surgery. This is because such individuals synthesise an atypical form of the enzyme serum cholinesterase which is much less effective in destroying the drug than the usual form of the enzyme, or because they fail to synthesise the enzyme in any significant amounts at all. The muscular and consequently respiratory paralysis induced by the drug is under these circumstances excessively prolonged. However unless exposed to this unusual and rather artificial feature of the environment, the drug suxamethonium, these people appear to be quite normal. Thus, both the rare genetical predisposition and the unusual environmental circumstance must occur together, before an abnormal clinical state appears. It is not surprising that such a disorder may not at first sight appear to be familial.

If we survey the range of human disease and abnormalities, we find at one extreme conditions such as sickle-cell anaemia, phenylketonuria, haemophilia, muscular dystrophy and so on – which we regard as inherited disorders because all individuals who have the appropriate genes develop the condition. At the other extreme are the typically environmental diseases – such as the severe infections like plague, anthrax, typhus – in which it appears that virtually everyone who is sufficiently exposed to the unfavourable environmental agent develops the disease. But in between there are many disorders, often common ones, in which both genetical and environmental factors are apparently important. Typical examples are schizophrenia, diabetes mellitus and peptic ulcer. In each of these conditions there is good evidence for genetical predisposition, and it is also clear that only a fraction of those genetically predisposed actually develop the condition. But we do not know the nature of the primary effects of the genes involved. Nor can we define very clearly the particular features of the environment which cause some individuals but not others among those genetically predisposed to develop the condition.

The elements of the situation can be illustrated by a simple diagram (fig. 9.1). The area encompassed by the outer circle represents a population, that enclosed with the inner circle those individuals in the population genetically predisposed to develop a particular kind of disease. The two lines drawn from the centre to the periphery divide the whole population into those people who happen to be exposed to environmental factors which tend

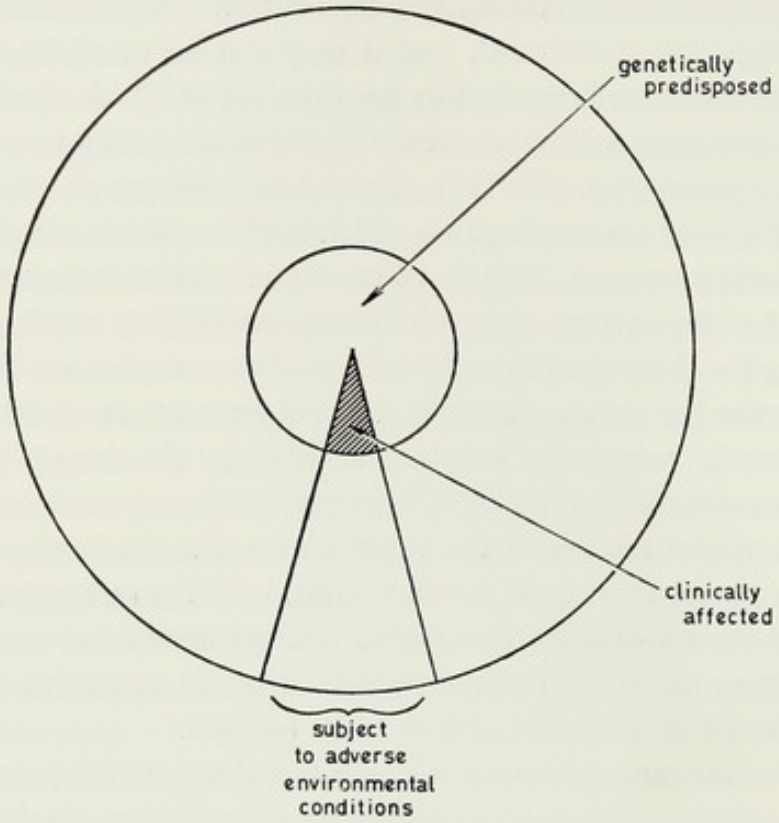


Fig. 9.1. Diagram illustrating in a simple form the roles of genetical predisposition and adverse environmental factors in the causation of disease (see text). Area bounded by outer circle represents a population, and the area enclosed within the inner circle those individuals in the population who are genetically predisposed to develop a particular kind of disease. The two lines from the centre to the periphery enclose a segment of the population subject to some particular environmental conditions that may elicit the disease. The hatched part of the segment represents those individuals who actually develop the disease.

to elicit the abnormality (in this case the smaller group), and those who are not so exposed. Only the small segment of the population who are both genetically predisposed and also subject to the unfavourable environmental situation, actually develop the clinical disorder.

The practical significance of this kind of formulation is that it emphasises how if by genetical research one could find ways of identifying the individuals in the population who are genetically predisposed to a particular condition, then one would be provided with a powerful tool with which to discover the critical environmental factors. In effect one would be able to ask what significant differences exist or have existed between the environmental circumstances of the predisposed individuals who actually develop the disease, and those who have not. And of course if these environmental circumstances can

be recognised, they can hopefully be adjusted in an appropriate manner for the genetically predisposed individuals before the clinical abnormality develops.

Clearly for any particular condition the fraction of the population genetically predisposed may be very small or quite large, and the proportion exposed to the adverse environmental situations can vary similarly. Furthermore one must expect that often many different genes can give rise in one way or another to a particular kind of predisposition, and since their effects will not be identical, the degree of predisposition will vary and be graded in severity. Similarly the relevant environmental factors are likely to vary in strength. So the lines shown in the diagram should be regarded as fuzzy rather than sharp.

Nevertheless this simple if diagrammatic approach provides a useful way of thinking about the problems posed by disease states, and particularly about those which are relatively common. It also leads to an interesting paradox. This is that the study of the genetics of many diseases may lead to their prevention or amelioration by purely environmental methods. Indeed it is very probable that one of the most important social and medical applications of genetical research will lie in the control of the environment, since the more it becomes possible to characterise the genetical constitution of an individual precisely, the more likely are we to see how to modify or tailor the environment according to his needs.

APPENDIX 1

Disorders due to specific enzyme deficiencies (inborn errors of metabolism)

A1.1 Disorders of carbohydrate metabolism

Enzyme deficiency	Condition
1. Hexokinase (red cell isozyme) [ATP : D-hexose 6-phosphotransferase] E.C. 2.7.1.1.	*Hexokinase deficiency haemolytic anaemia* Defective phosphorylation of glucose (fig. A1) limits glycolysis and results in premature breakdown of red cells with chronic haemolytic anaemia. No hexokinase deficiency in leucocytes. Ref.: Valentine et al. (1967).
2. Phosphohexose isomerase (glucose phosphate isomerase) [D-glucose-6-phosphate ketol-isomerase] E.C. 5.3.1.9	*Phosphohexoseisomerase (glucose phosphate isomerase deficiency)* Defective conversion of glucose-6-phosphate to fructose-6-phosphate (fig. A1) in red cells limits glycolysis and results in severe haemolytic anaemia. Enzyme abnormality also present in leucocytes and probably other tissues but this causes no obvious clinical consequences. Several different alleles producing structurally altered forms of the enzyme with deficient functional activity occur, as well as a number of other variants. Refs.: Baughan et al. (1968), Paglia et al. (1969), Detter et al. (1968).
3. Phosphofructokinase (muscle type isozyme) [ATP : D-fructose-6-phosphate 1-phosphotransferase] E.C. 2.7.1.11	*Phosphofructokinase deficiency* (p. 165) Restriction of glycolysis in muscle due to gross reduction of the enzyme activity leads to glycogen accumulation. Muscular weakness and stiffness develops on prolonged exertion. Partial reduction of activity also present in red cells and causes chronic mild haemolysis. The residual phosphofructokinase activity in red cells (about 50%) is due to a different isozyme not

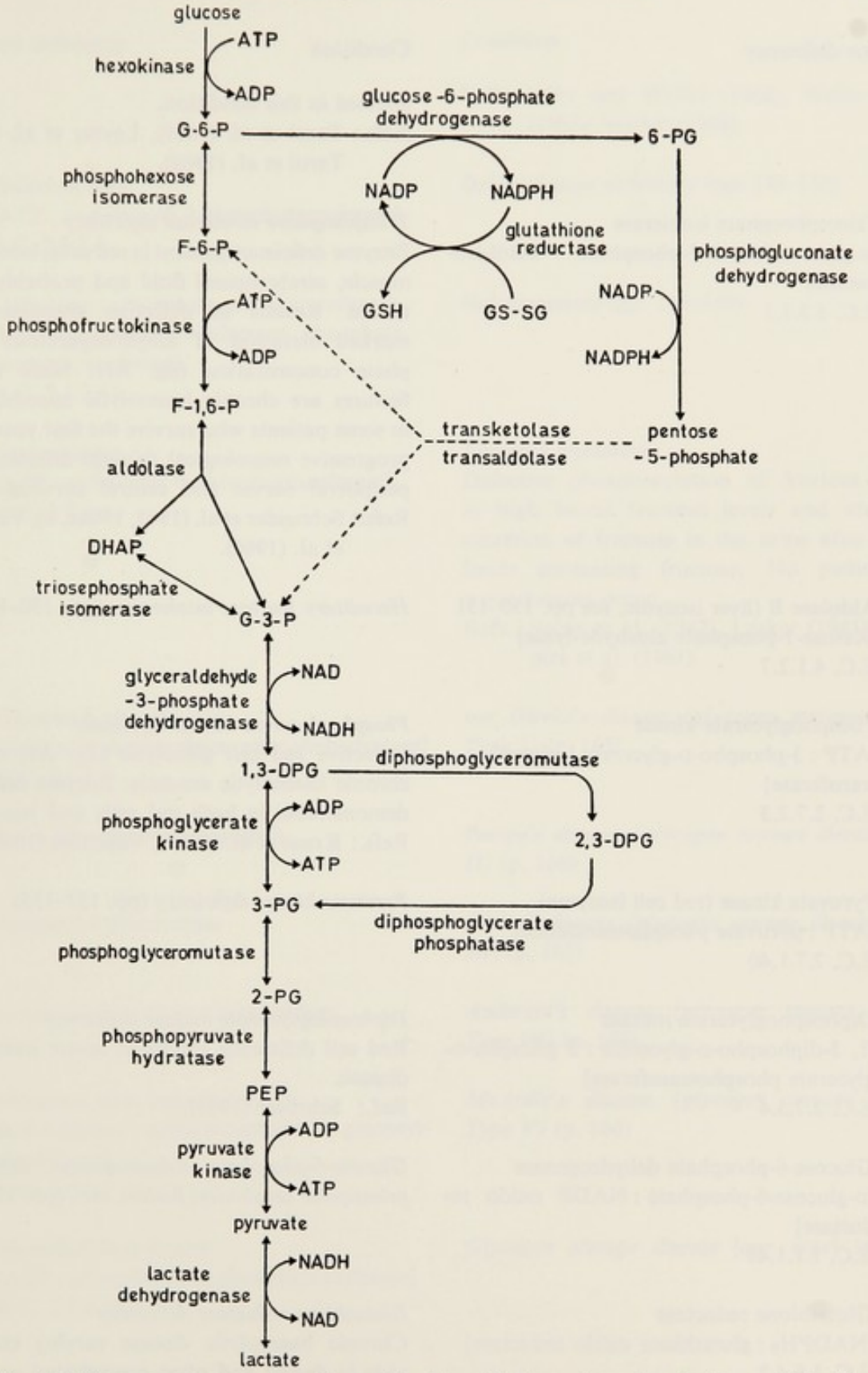


Fig. A1. Enzymes concerned in the metabolism of glucose in red cells (from Valentine 1968).

Enzyme deficiency	Condition
	affected in this condition. Refs.: Tarui et al. (1965), Layzer et al. (1967), Tarui et al. (1969).
4. Triosephosphate isomerase [D-glyceraldehyde-3-phosphate ketol-isomerase] E.C. 5.3.1.1	*Triosphosphate isomerase deficiency* Enzyme deficiency present in red cells, leucocytes, muscle, cerebrospinal fluid and probably other tissues. Results in defective glycolysis with marked elevation of dihydroxyacetone phosphate concentration (fig. A1). Main clinical features are chronic haemolytic anaemia, and in some patients who survive the first year or so, progressive neurological damage affecting both peripheral nerves and central nervous tissue. Refs.: Schneider et al. (1965, 1968a, b), Valentine et al. (1966).
5. Aldolase B (liver isozyme, see pp. 150–151 [Ketose-1-phosphate aldehyde-lyase] E.C. 4.1.2.7	*Hereditary fructose intolerance* (pp. 150–154)
6. Phosphoglycerate kinase [ATP : 3-phospho-D-glycerate 1-phosphotransferase] E.C. 2.7.2.3	*Phosphoglycerate kinase deficiency* Defective red cell glycolysis (fig. A1) causing chronic haemolytic anaemia. Enzyme deficiency demonstrable in both red cells and leucocytes. Refs.: Kraus et al. (1968), Valentine (1968).
7. Pyruvate kinase (red cell isozyme) [ATP : pyruvate phosphotransferase] E.C. 2.7.1.40	*Pyruvate kinase deficiency* (pp. 154–155)
8. Diphosphoglycerate mutase [1, 3-diphospho-D-glycerate : 3 phospho-D-glycerate phosphotransferase] E.C. 2.7.5.4	*Diphosphoglycerate mutase deficiency* Red cell deficiency results in severe haemolytic disease. Ref.: Schröter (1965).
9. Glucose-6-phosphate dehydrogenase [D-glucose-6-phosphate : NADP oxido reductase] E.C. 1.1.1.49	*Glucose-6-phosphate dehydrogenase deficiency, primaquine sensitivity, favism, etc.* (pp. 121–134)
10. Glutathione reductase [$NADPH_2$: glutathione oxido reductase] E.C. 1.6.4.2	*Glutathione reductase deficiency* Chronic haemolytic disease varying considerably in degree and often precipitated or accentuated by drugs.

Enzyme deficiency	Condition
	Refs.: Löhr and Waller (1962), Waller et al. (1965), Waller (1968).
11. Galactokinase [ATP : D-galactose 1-phosphotransferase] E.C. 2.7.1.6	*Galactokinase deficiency* (pp. 149–150)
12. Galactose-1-phosphate uridyl transferase [UDP glucose : α-D-galactose phosphate uridylyltransferase] E.C. 2.7.7.12	*Galactosaemia* (pp. 148–149)
13. Fructokinase [ATP : D-fructose 6-phosphotransferase] E.C. 2.7.1.4	*Essential fructosuria* Defective phosphorylation of fructose results in high blood fructose levels and abnormal excretion of fructose in the urine after taking foods containing fructose. No pathological consequences occur. Refs.: Sachs et al. (1942), Lasker (1941), Schapira et al. (1961).
14. Glucose-6-phosphatase [D-glucose 6-phosphate phosphohydrolase] E.C. 3.1.3.9	*von Gierke's disease (glycogen storage disease Type I)* (p. 165)
15. α-1,4 glucosidase (lysosomal)	*Pompe's disease (glycogen storage disease Type II)* (p. 166)
16. Amylo-1,6-glucosidase	*Forbes' disease (glycogen storage disease, Type III)* (p. 165)
17. Amylo-(1,4→1,6)-transglucosidase	*Andersen's disease (glycogen storage disease Type IV)* (p. 164)
18. Phosphorylase (muscle type) [α-1,4-glucan : orthophosphate glucosyltransferase] E.C. 2.4.1.1	*McArdle's disease (glycogen storage disease Type V)* (p. 164)
19. Phosphorylase kinase [ATP : phosphorylase phosphotransferase] E.C. 2.7.1.38	*Glycogen storage disease* [*one type*] (p. 165)

Enzyme deficiency	Condition
20. Glycogen synthetase [UDP glucose : glycogen α-4-glucosyltransferase] E.C. 2.4.1.11	*Glycogen synthetase deficiency* (pp. 163–164)
21. L-xylulose reductase [Xylitol : NADP oxidoreductase (L-xylulose forming)] E.C. 1.1.1.10	*Congenital pentosuria* L-xylulose (L-xyloketose) is a normal intermediate in the pathway by which D-glucuronic acid is metabolised (fig. A2). Deficiency of the reductase results in accumulation of L-xylulose and its continuous excretion in large amounts in the urine. Increased excretion of L-arabitol has also been found. No pathological consequences occur. The urinary output of L-xylulose is greatly enhanced if glucuronic acid is given, or if certain drugs which stimulate glucuronic acid formation and are themselves excreted as glucuronides are taken. Refs.: Enklewitz and Lasker (1935), Lasker et al. (1936), Touster (1959), Hiatt (1966).
22. Isomaltase (maltase 1a) and sucrase (maltase 1b)	*Sucrose and isomaltose intolerance* Inability to hydrolyse sucrose and isomaltose (formed from starch) in the course of intestinal digestion. On a normal diet there is a persistent chronic diarrhoea with frothy and liquid acid stools. The low *p*H of the stools is due to the presence of lactic and other organic acids formed from undigested carbohydrate by bacterial fermentation in the large intestine. Refs.: Weijers et al. (1960), Prader and Auricchio (1965), Rey and Frézal (1967).
23. Lactase	*Congenital lactose intolerance* Failure to hydrolyse ingested lactose in milk. Chronic diarrhoea starting a few days after birth results. Refs.: Holzel et al. (1959), Lifschitz (1966).

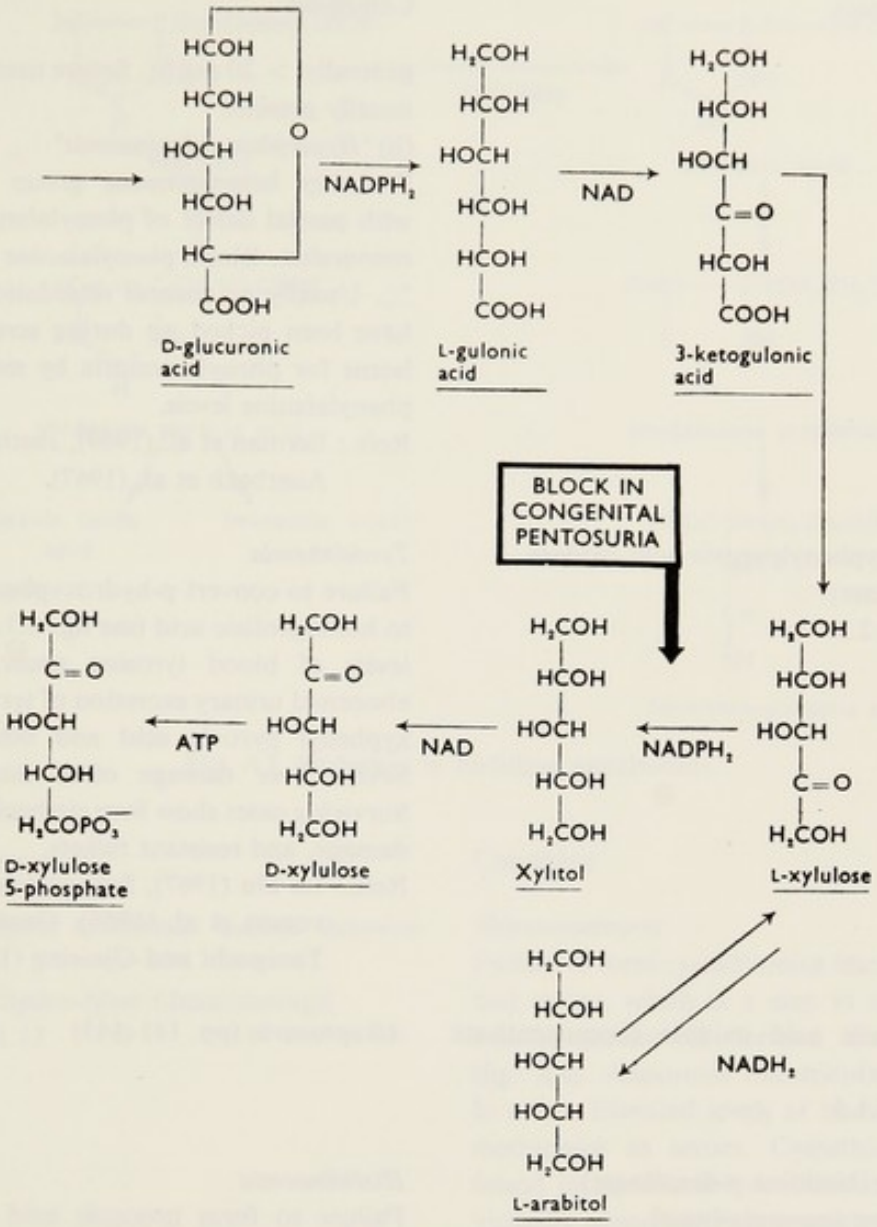


Fig. A2. The glucuronic acid oxidation pathway showing the site of the metabolic block in congenital pentosuria.

A1.2 Disorders of aminoacid metabolism

Enzyme deficiency	Condition
1. Phenylalanine 4-hydroxylase E.C. 1.14.3.1	(a) *Phenylketonuria* (pp. 145–148). Complete or almost complete failure to convert phenylalanine to tyrosine (see fig. 6.2, p. 145). Blood phenylalanine levels in untreated state

Enzyme deficiency	Condition
	generally > 20 mg %. Severe mental retardation usually present. (b) '*Hyperphenylalaninaemia*' Probably heterogeneous group of conditions with partial defect of phenylalanine to tyrosine conversion. Blood phenylalanine levels 6-20 mg %. Usually no mental retardation. Most cases have been picked up during screening of newborns for phenylketonuria by measuring blood phenylalanine levels. Refs.: Berman et al. (1969), Justice et al. (1967), Auerbach et al. (1967).
2. p-Hydroxyphenylpyruvic acid oxidase (hydroxylase) E.C. 1.14.2.2	*Tyrosinaemia* Failure to convert p-hydroxyphenylpyruvic acid to homogentisic acid (see fig. 6.1, p. 142). High levels of blood tyrosine occur, with grossly abnormal urinary excretion of tyrosine, p-hydroxyphenyl pyruvic acid and other derivatives. Severe liver damage often fatal in infancy. Surviving cases show liver cirrhosis, renal tubular damage, and resistant rickets. Refs.: La Du (1967), Scriver et al. (1967), Halvorsen et al. (1966), Gentz et al. (1965), Taniguchi and Gjessing (1965).
3. Homogentic acid oxidase (homogentisate oxygenase) E.C. 1.13.1.5	*Alkaptonuria* (pp. 141–143)
4. Histidase (histidine α-deaminase) [L-histidine ammonia-lyase] E.C. 4.3.1.3	*Histidinaemia* Failure to form urocanic acid from histidine (fig. A3). Elevated blood histidine. Increased urinary excretion of histidine imidazole-pyruvic acid. Absence of formininoglutamic acid in urine after histidine load. Absence of urocanic acid in sweat. Delayed onset of speech, and speech defects common. About half the patients some degree of mental retardation. Refs.: Ghadimi et al. (1961), La Du et al. (1962), Auerbach et al. (1962), Ghadimi and Partington (1967), La Du (1967).

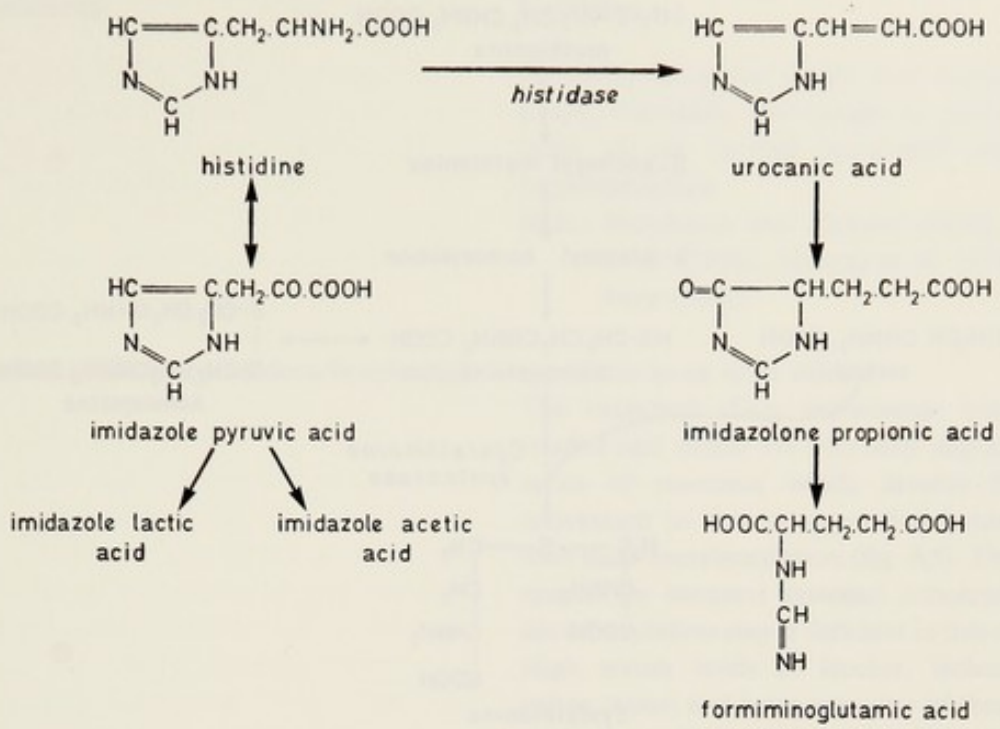


Fig. A3. Pathways in histidine metabolism.

Enzyme deficiency	Condition
5. Cystathionine synthetase (serine dehydratase) [L-serine hydro-lyase (deaminating)] E.C. 4.2.1.13	*Homocystinuria* Failure to form cystathionine from homocysteine and serine, which is a step in the pathway by which cysteine is formed from methionine (fig. A4). Abnormal excretion of homocystine in urine. Elevated levels of homocystine and of methionine in serum. Cystathionine normally found in significant quantities in brain tissue is virtually absent in this condition. Clinical syndrome includes mental retardation, ectopia lentis, skeletal abnormalities, and thromboembolic phenomena in both arteries and veins. Refs.: Gerritsen and Waisman (1964a, b), Carson et al. (1965), Brenton et al. (1966), Uhlendorf and Mudd (1968).
6. Cystathionase (homoserine dehydratase) [L-homoserine hydro-lyase (deaminating)] E.C. 4.2.1.15	*Cystathioninuria* Failure to cleave cystathionine to give cystine, α-ketobutyrate and ammonia. This reaction is a step in the major pathway of methionine metabolism (fig. A4). There is a high urinary excretion of cystathionine, and an increased concentration

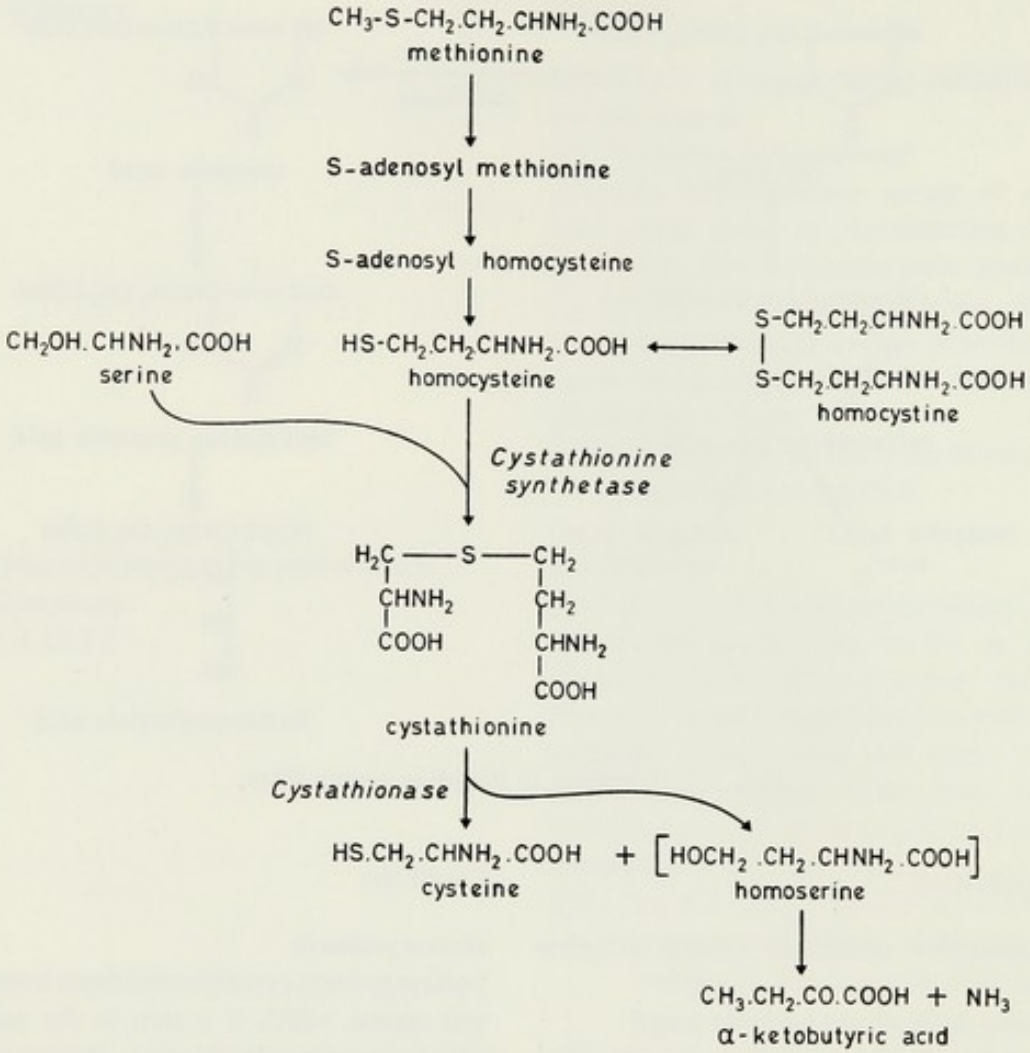


Fig. A4. Pathway for the formation of cysteine from methionine.

Enzyme deficiency	Condition
	of cystathionine in the tissues and in serum. Mental retardation and psychotic abnormalities have been seen in different patients, but it is unclear whether this is a consequence of the metabolic disorder.
	Refs.: Harris et al. (1959), Frimpter et al. (1963), Frimpter (1965, 1967), Finkelstein et al. (1966).
7. Iodotyrosine deiodinase	*Goitrous cretinism* (one type)
	Defective deiodination of mono- and di-iodotyrosine during thyroid hormone synthesis. Continuous leakage of these hormone precursors from the thyroid and from the body leads to

Enzyme deficiency	Condition
	depletion of iodine stores and thyroid hyperplasia. Clinically this results in gross enlargement of the thyroid associated with severe hypothyroidism. Refs.: Hutchison and McGirr (1956), Querido et al. (1956), Murray et al. (1965), Stanbury (1966).
8. 'Branched chain ketoacid decarboxylase(s)'	(a) *Maple syrup urine disease* The branched chain aminoacids leucine, isoleucine and valine are normally degraded via a series of reactions which involve first their conversion to the corresponding ketoacids and then their decarboxylation (fig. A5). The enzyme system (or separate systems) concerned in the decarboxylation step is deficient in this condition. High serum levels of leucine, isoleucine and valine occur; and large amounts of these aminoacids as well as the corresponding ketoacids (isovaleric acid, α-methylbutyric acid and isobutyric acid) are excreted in the urine which has a characteristic smell. Alloisoleucine probably derived from isoleucine is also present in ab-

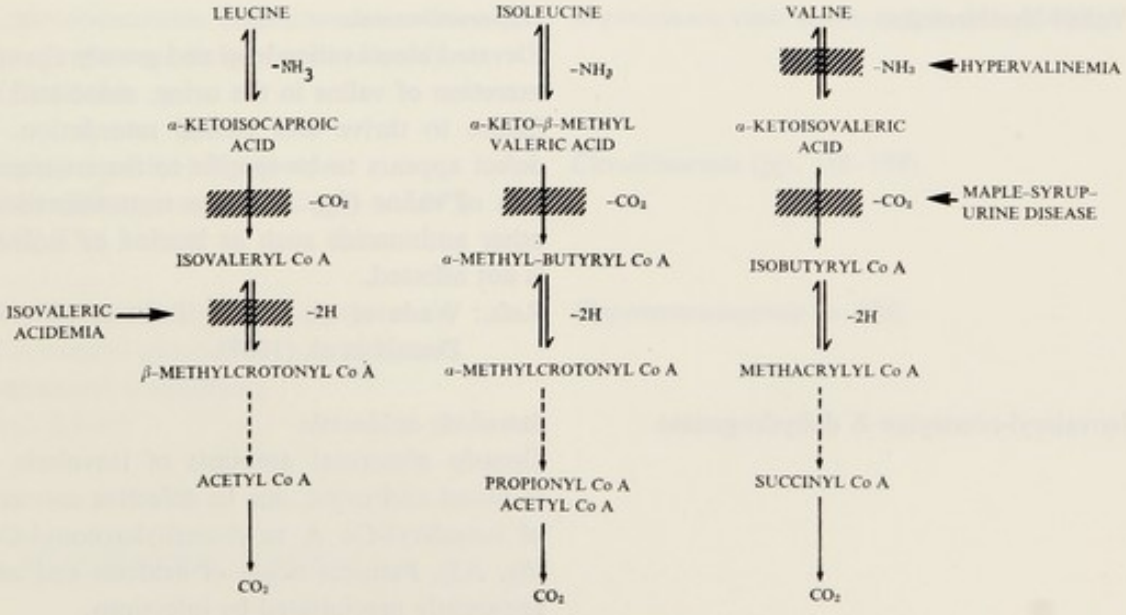


Fig. A5. Pathways in the metabolism of the branched chain aminoacids. Sites of metabolic blocks in maple syrup urine disease, hypervalinaemia and isovaleric acidaemia. (From Budd et al. 1967.)

Enzyme deficiency	Condition
	normal amounts in the serum. Rapidly progressive neurological disease with marked cerebral degeneration is usually apparent shortly after birth, and is generally fatal within weeks or months. Refs.: Dancis et al. (1959), Mackenzie and Woolf (1959), Dancis et al. (1963), Snyderman (1967), Goedde and Keller (1967). (b) *Intermittent branched-chain ketonuria* Slightly less severe deficiency of the enzyme(s) which are grossly deficient in maple syrup urine disease. Intermittent episodes with increased excretion of branched chain ketoacids and aminoacids occur in urine which may also have the characteristic maple syrup odour. During the episodes, signs and symptoms attributable to neurological disorder (toxic encephalopathy) occur and may be fatal. Between episodes there may be no clinical abnormality, and the urinary excretion of ketoacids and aminoacids is within the normal range. Episodes apparently precipitated by infections or high protein intake. Refs.: Kiil and Rokkenes (1964), Morris et al. (1966), Dancis et al. (1967).
9. Valine transaminase	*Hypervalinaemia* Elevated blood valine level and grossly abnormal excretion of valine in the urine, associated with failure to thrive and mental retardation. The defect appears to be specific to the transamination of valine (fig. A5). The transamination of other aminoacids such as leucine or isoleucine is not affected. Refs.: Wada et al. (1963), Tada et al. (1967), Dancis et al. (1967).
10. Isovaleryl-coenzyme A dehydrogenase	*Isovaleric acidaemia* Grossly abnormal amounts of isovaleric acid in blood and urine, due to defective conversion of isovaleryl-Co A to β-methylcrotonyl-Co A (fig. A5). Periodic bouts of acidosis and coma, apparently precipitated by infections. Refs.: Tanaka et al. (1966), Budd et al. (1967).

Enzyme deficiency	Condition
11. Proline oxidase [L-proline : NAD(P) 5-oxidoreductase) E.C. 1.5.1.2	*Hyperprolinaemia* Abnormal elevation of serum proline due to failure of conversion to Δ'-pyrroline-5-carboxylate. Oxidation of hydroxyproline is unaffected. However there is an increased urinary excretion not only of proline but also hydroxyproline and glycine due probably to saturation of the common transport system for these aminoacids in the renal tubules by excess proline (see p. 182). Renal abnormalities and mental retardation have been found in some cases, but how far they are a consequence of the specific metabolic disorder is unclear. Refs.: Shafer et al. (1962), Efron (1965).
12. Hydroxyproline oxidase	*Hydroxyprolinaemia* Abnormal amounts of hydroxyproline in serum and urine due to failure of conversion to Δ'-pyrroline-3-hydroxy-5-carboxylate. No abnormality in collagen metabolism or in peptide-bound urinary hydroxy proline. Proline metabolism apparently normal. Severe mental retardation associated with the defect. Refs.: Efron et al. (1965).
13. Argininosuccinase (argininosuccinate lyase) [L-argininininosuccinate arginine-lyase] E.C. 4.3.2.1	*Argininininosuccinic aciduria* (pp. 156–158)
14. Argininosuccinate synthetase [L-citrulline : L-aspartate ligase (AMP)] E.C. 6.3.4.5	*Citrullinaemia* (pp. 158–159)
15. Ornithine carbamoyl transferase [Carbamoyl phosphate : L-ornithine carbamoyl transferase] E.C. 2.1.3.3	*Hyperammonaemia* (p. 158)

A1.3 Miscellaneous disorders

1. Xanthine oxidase [Xanthine : O_2 oxidoreductase] E.C. 1.2.3.2.	*Xanthinuria* There is a failure to form uric acid from xanthine so that xanthine effectively replaces uric

Enzyme deficiency	Condition
	acid as the end product of purine metabolism. Serum and urinary uric acid levels are extremely low, and there is a grossly abnormal excretion of xanthine in the urine. Due to the low solubility of xanthine, xanthine calculi tend to form in the renal tract. Refs.: Dent and Philipot (1954), Watts et al. (1963), Engleman et al. (1964).
2. Reduced nicotinamide-adenine dinucleotide (NADH) oxidase	*Chronic granulomatous disease* Deficiency of this enzyme in polymorphonuclear leucocytes appears to render them incapable of killing many types of bacteria which they can ingest normally. Affected individuals suffer from persistent chronic infections characterised by suppuration and granulomatous lymphadenitis, granulomatous infiltrations especially of the lungs, and hepatosplenomegaly. Organisms involved are usually of low grade virulence, e.g. staphylococci, *Aerobacter aerogenes* and *Klebsiella sp.* Refs.: Holmes et al. (1966), Baehner and Karnovsky (1968).
3. Sulphite oxidase [Sulphite : O_2 oxidoreductase] E.C. 1.8.3.1	*Sulphite oxidase deficiency* Inability to convert sulphite to sulphate leading to increased formation of S-sulpho-L-cysteine and thiosulphate. Greatly increased amounts of S-sulpho-L-cysteine, sulphite and thiosulphate in the urine, but virtually no inorganic sulphate present. Progressive neurological abnormalities, dislocation of lenses, and mental retardation. Ref.: Mudd et al. (1967).
4. Catalase [H_2O_2 : H_2O_2 oxidoreductase] E.C. 1.11.1.6	*Acatalasia* Gross deficiency of catalase in all tissues. In some individuals ulceration of the mucosae of the nose and mouth occurs and severe oral gangrene may develop. But other individuals with less than 1% of the normal catalase level remain quite healthy and apparently suffer no ill effects. Refs.: Takahara (1952), Kaziro et al. (1952), Aebi et al. (1964), Aebi (1967), Takahara (1968), Aebi et al. (1968).

Enzyme deficiency	Condition
5. Glutathione peroxidase (Glutathione : H_2O_2 oxidoreductase)	*Red cell glutathione peroxidase deficiency* Mild chronic haemolytic disease. Ref.: Necheles et al. (1969).
6. Methaemoglobin reductase (DPNH : methaemoglobin oxidoreductase)	*Congenital methaemoglobinaemia* (p. 254)
7. Orotidine-5′-phosphate pyrophosphorylase [Orotidine-5′-phosphate : pyrophosphate phosphoribosyltransferase] E.C. 2.4.2.10 *and* Orotidine-5′-phosphate decarboxylase [Orotidine-5′-phosphate carboxylyase] E.C. 4.1.1.23	*Orotic aciduria* A gross deficiency of *two* sequential enzymes, which are concerned in the conversion of orotic acid to uridine-5′-phosphate (fig. A6), is present. There is marked retardation of growth and development, severe megaloblastic anaemia, and abnormal excretion of orotic acid in the urine. Refs.: Huguley et al. (1959), Smith et al. (1961), Fallon et al. (1964), Krooth (1964), Howell et al. (1967).
8. Glucocerebrosidase	*Gaucher's disease* (pp. 166–169)
9. Arylsulphatase-A [Aryl-sulphate sulphohydrolase] E.C. 3.1.6.1	*Metachromatic leucodystrophy* (p. 168) Accumulation of ceramide-galactose-3-sulphate with progressive degeneration of the central nervous system. Refs.: Austin et al. (1963), Mehl and Jatzkewitz (1965), Jatzkewitz and Mehl (1969).

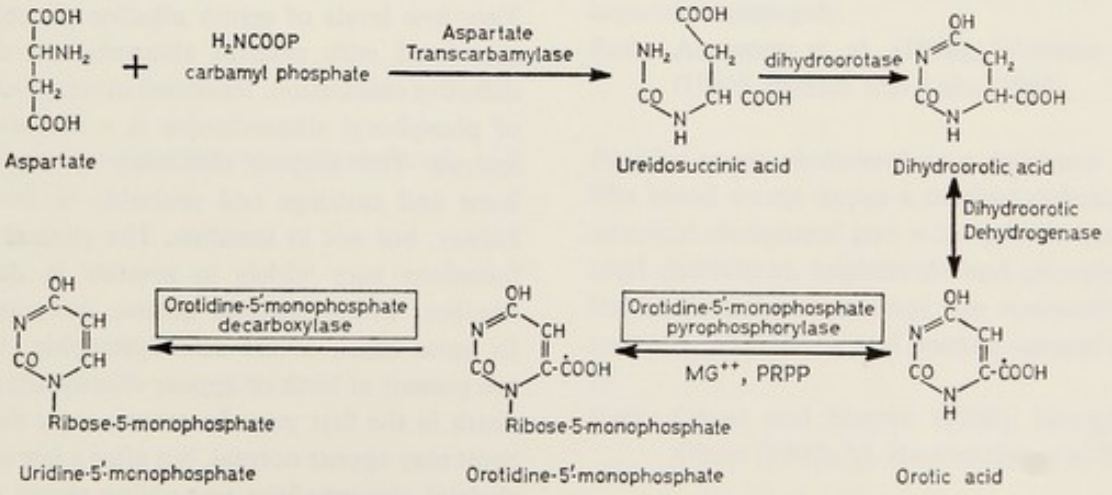


Fig. A6. Pathway of biosynthesis of uridine-5′-monophosphate showing sites of enzyme defects in orotic aciduria. (From Wuu and Krooth 1968.)

Enzyme deficiency	Condition
10. Ceramide trihexosidase	*Fabry's disease (angiokeratoma corporis diffusum)* (p. 168) Accumulation of ceramide trihexoside. Development of skin lesions (small dark purple macules and papules), ocular abnormalities (corneal opacities, cataracts and retinal oedema), and cardiovascular, neurological and gastrointestinal disorders. Severe burning pains in the extremities are a characteristic feature. Inheritance X linked. Refs.: Wise et al. (1962), Sweeley and Klionsky (1963), Opitz et al. (1965), Brady et al. (1967a, b).
11. Sphingomyelinase	*Niemann–Pick disease* (p. 168) Accumulation of sphingomyelin. Progressive degeneration of central nervous system associated with hepatosplenomegaly. Refs.: Fredrickson (1966), Brady et al. (1966b), Sloan et al. (1969).
12. 'Acid' β-galactosidase	*Generalised gangliosidosis* (pp. 168, 170)
13. α-fucosidase	*Fucosidosis* (p. 168)
14. Serum cholinesterase (pseudocholinesterase) [Acylcholine acyl-hydrolase] E.C. 3.1.1.8	*Suxamethonium sensitivity* (pp. 109–119)
15. Alkaline phosphatase	*Hypophosphatasia* Very low levels of serum alkaline phosphatase associated with skeletal abnormalities due to defective ossification. Increased urinary excretion of phosphoryl ethanolamine is a characteristic feature. The enzyme deficiency is present in bone and cartilage and probably in liver and kidney, but not in intestine. The clinical manifestations vary widely in severity in different families, which suggests genetic heterogeneity. In some cases, severe osteodystrophic changes are present at birth or appear shortly after, with death in the first year. In others, early development may appear normal, but after a few months skeletal abnormalities resembling severe rickets develop. The teeth are hypoplastic and shed prematurely. In still other cases, the disorder is

Enzyme deficiency	Condition
	not recognised until adult life, the patients presenting with bony deformities and spontaneous fractures, though there is often a history of rickets in childhood. Refs.: Rathbun (1948), Sobel et al. (1953), Fraser (1957), Currarino et al. (1957), Harris and Robson (1959), Rathbun et al. (1961), Danovitch et al. (1968), Rasmussen (1968).
16. Adenosine triphosphatase	*Red cell ATPase deficiency* Haemolytic anaemia occurring in apparently heterozygous individuals. Ref.: Harvald et al. (1964).
17. Pancreatic lipase [Glycerol ester hydrolase] E.C. 3.1.1.3	*Congenital pancreatic lipase deficiency* Lipase in intestinal juice reduced to about 10% of normal, resulting in marked inability to digest fats. About 50-80% of ingested fats are excreted in the stools which are very bulky and oily. Refs.: Sheldon (1964), Rey et al. (1966).
17. Acid lipase	*Wolman's disease* Deficiency of acid lipase probably of lysosomal origin associated with widespread occurrence in visceral organs of 'foamy' cells containing large amounts of triglycerides and cholesterylesters. The condition is characterised by xanthomatosis, calcification of the adrenals, and hepatosplenomegaly. Refs.: Abramov et al. (1956), Wolman et al. (1961), Patrick and Lake (1969).
18. Lecithin : cholesterol acyltransferase	*Familial serum cholesterol ester deficiency* The blood serum shows a marked deficiency of esterified cholesterol and α-lipoprotein but high total cholesterol, triglyceride and phospholipid levels. Associated findings are normochromic anaemia, proteinuria and marked corneal opacity. Refs.: Gjone and Norum (1968), Norum and Gjone (1967a,b), Hamnström et al. (1969).

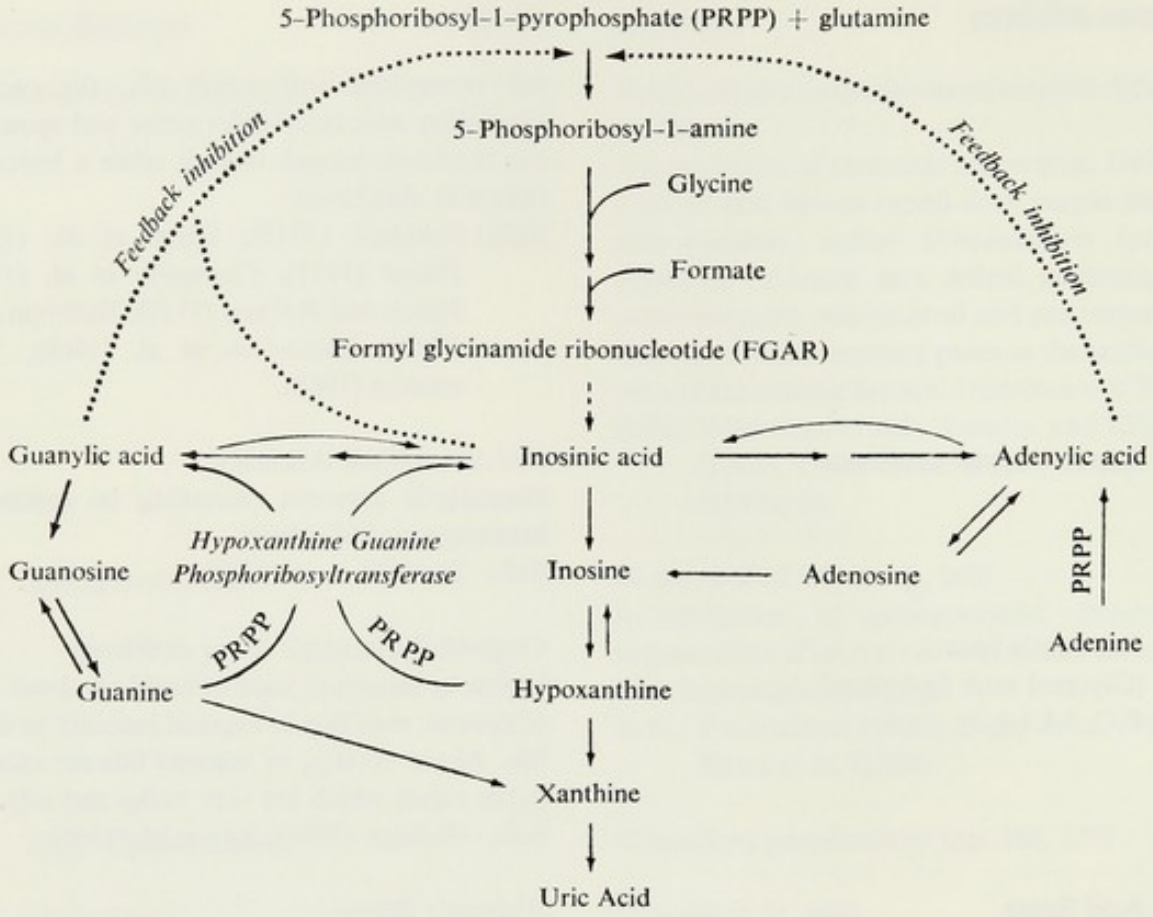


Fig. A7. Pathways in purine biosynthesis and interconversions (Seegmiller et al. 1967).

Enzyme deficiency	Condition
19. Hypoxanthine-guanine phosphoribosyl transferase	(a) *Lesch–Nyhan syndrome* (pp. 257–258) (b) *Gout due to uric acid overproduction* pp. 257–258 The role the enzyme is thought to play in the pathways of purine biosynthesis is indicated in fig. A7. It is not however known why the specific deficiency of the enzyme should result in over-production of uric acid which occurs in these syndromes.
21. 2-oxo-glutarate : glyoxalate carboligase	*Primary hyperoxaluria* (one type) Increased urinary excretion of oxalate, glycolate and glyoxalate. Calcium oxalate calculi form in urinary tract, and calcium oxalate may be deposited within renal parenchyma (nephro-calcinosis) and more widely throughout the body (oxalosis). Refs.: Archer et al. (1957), Hockaday et al. (1964), Koch et al. (1967).

Enzyme deficiency	Condition
22. D-glyceric dehydrogenase	*Primary hyperoxaluria* (one type) Increased urinary excretion of oxalate and L-glyceric acid. Oxalate calculi form in renal tract. Ref.: Williams and Smith (1968).
23. Trypsinogen	*Trypsinogen deficiency disease* The duodenal juice lacks not only trypsin activity but also chymotrypsin and carboxypeptidase activity. This is because trypsin, formed from trypsinogen, is required to activate the chymotrypsinogen and procarboxypeptidase which are formed in normal amounts. Consequently no protein digestion occurs in the duodenum and small intestine. The affected newborn fails to grow and there is a progressive development of marked hypoproteinaemia, oedema, and anaemia. The condition responds well to the addition of hydrolysed protein to the diet. Refs.: Townes (1965), Townes et al. (1967).
24. Enterokinase (enteropeptidase) E.C. 3.4.4.8	*Intestinal enterokinase deficiency* Deficiency of enterokinase results in a failure to convert trypsinogen, which is present in normal amounts, to trypsin. In the absence of trypsin activity, chymotrypsinogen and procarboxypeptidase are not converted to chymotrypsin and carboxypeptidase. So there is a failure of protein digestion, with consequences identical to those seen in trypsinogen deficiency disease (see above). Ref.: Hadorn et al. (1969).
25. Methylmalonyl-Co A carbonyl mutase [Methylmalonyl-Co-A Co A - Carbonyl mutase] E.C. 5.4.99.2	*Methylmalonic acidemia* Methylmalonyl-Co A carbonylmutase converts L-methylmalonyl-Co A to succinyl-Co A and requires the cobamide coenzyme form of Vit B_{12}. Its deficiency results in a failure in the metabolic pathway by which succinate is formed from propionate. There is a massive urinary excretion of methylmalonic acid. The children fail to thrive and have a severe ketoacidosis. They are retarded and often die in early life. Two distinct types of defect have been recognised. One responds well to massive doses of Vit B_{12}, while the other is quite unresponsive to this treatment. Refs.: Oberholzer et al. (1967), Rosenberg et al. (1968), Morrow et al. (1969).

APPENDIX 2

Enzyme and protein polymorphisms*

Enzyme or protein	Method of detection	References
1. Haemoglobin	Electrophoresis	See pp. 5–12, 213–214
2. Haptoglobin	Electrophoresis	See pp. 67–75, 221–224
3. Transferrin	Electrophoresis	Smithies (1957), Smithies and Hiller (1959), Giblett (1962).
4. Serum α-globulin (*Gc*)	Immuno-electrophoresis	Hirschfeld (1959), Hirschfeld (1962), Bearn et al. (1964), Rienskou (1968).
5. Immunoglobulins Heavy chains (*Gm* system)	Immunological (antibody neutralisation)	Grubb (1956), Grubb and Laurell (1956), Steinberg (1967).
6. Immunoglobulins Light chains (*Inv* system)	Immunological (antibody neutralisation)	Ropartz et al. (1961), Ritter and Wendt (1964), Steinberg (1967).
7. Serum β-lipoproteins (*Ag* system)	Immuno-diffusion and precipitation	Allison and Blumberg (1961), Blumberg et al. (1962), Hirschfeld and Okachi (1967), Morganti et al. (1967), Hirschfeld (1968).
8. Serum β-lipoproteins (*Lp* system)	Immuno-diffusion and precipitation	Berg (1963, 1965, 1966, 1968), Bundschuh and Vogt (1965), Bütler (1967).
9. Caeruloplasmin	Electrophoresis	Shreffler et al. (1967).
10. Serum α_1 trypsin inhibitor (*Pi* system)	Electrophoresis and immuno-electrophoresis	Fagerhol and Braend (1965), Fagerhol and Laurell (1967), Fagerhol (1967, 1968, 1969)
11. Third component of complement (*C'3* system)	Electrophoresis	Alper and Propp (1968), Azen and Smithies (1968).
12. Serum α_2-macroglobulin (*Xm* system)	Immuno-diffusion and precipitation	Berg and Bearn (1966, 1968).

* Situations in which two or more alleles, each producing a discrete effect on the formation of a protein, have been found with frequencies greater than 0.01 in one or several major human populations.

Enzyme or protein	Method of detection	References
13. Serum α_1-acid glycoprotein (oroseromucoid)	Electrophoresis	Schmid et al. (1964, 1965).
14. Glucose-6-phosphate dehydrogenase [D-glucose-6-phosphate: NADP oxidoreductase] E.C. 1.1.1.49	Quantitative assay and electrophoresis	See pp. 121–131, 220–221
15. Phosphoglucomutase (Locus PGM_1) [α-D-glucose-1, 6-diphosphate: α-D-glucose-1-phosphate phosphotransferase] E.C. 2.7.5.1	Electrophoresis	See pp. 46–53, 224
16. Phosphoglucomutase (Locus PGM_3)	Electrophoresis	See pp. 50–53, 224
17. Placental alkaline phosphatase	Electrophoresis	See pp. 30–32
18. Serum cholinesterase (Locus E_1) [acylcholine acyl-hydrolase] E.C. 3.1.1.8	Quantitative inhibition tests	See pp. 109–119
19. Serum cholinesterase (Locus E_2)	Electrophoresis	See pp. 119–120
20. Peptidase A (dipeptidase)	Electrophoresis	See pp. 27–29
21. Liver acetyl-transferase	Metabolic loading test (isoniazid)	See pp.187–190
22. Red cell acid phosphatase	Electrophoresis	See pp. 134–140
23. Adenylate kinase [ATP : AMP phosphotransferase] E.C. 2.7.4.3	Electrophoresis	Fildes and Harris (1966), Rapley et al. (1967), Bowman et al. (1967).
24. Adenosine deaminase [Adenosine aminohydrolase] E.C. 3.5.4.4	Electrophoresis	Spencer et al. (1968).
25. Phosphogluconate dehydrogenase [6-phospho-D-gluconate : NADP oxidoreductase (decarboxylating)] E.C. 1.1.1.44	Electrophoresis	Fildes and Parr (1963), Parr (1966), Carter et al. (1968).
26. Erythrocyte nicotinamide adenine dinucleotide nucleosidase (NADase) [NAD glycohydrolase] E.C. 3.2.2.5	Quantitative assay	Ng et al. (1968).
27. Galactose-1-phosphate uridyl transferase (Duarte variant) [UDP glucose : α-D-galactose-1-phosphate uridylyl transferase] E.C. 2.7.7.12	Quantitative assay and electrophoresis	Beutler et al. (1965, 1966), Mathai and Beutler (1966).
28. Glutathione reductase	Electrophoresis	Long (1967).
29. Pancreatic amylase	Electrophoresis	Kamaryt and Laxova (1965)
30. Peptidase D (prolidase)	Electrophoresis	Lewis and Harris (1969b).

References

ABRAMOV, A., SCHORR, S. and WOLMAN, M. (1956). Fatal exfoliative dermatitis following injection of triple antigen and Salk vaccine. Am. J. Diseases Children *91*, 282.

AEBI, H. (1967). The investigation of inherited enzyme deficiencies with special reference to acatalasia. Proc. Third Int. Congr. Human Genetics. Johns Hopkins Press, Baltimore, p. 189.

AEBI, H., BAGGIOLINI, M., DEWALD, B., LAUBER, E., SUTER, H., MICHELI, A. and FREI, J. (1964). Observations in two Swiss families with acatalasia. Enzymol. Biol. Clin. *4*, 121.

AEBI, H., BOSSI, E., CANTZ, M., MATSUBARA, S. and SUTER, H. (1968). Acatalas(em)ia in Switzerland. *In:* 'Hereditary disorders of erythrocyte metabolism', ed. E. Beutler. Grune & Stratton, New York.

AGER, J. A. M. and LEHMANN, H. (1958). Observations on some 'fast' haemoglobins: K, J, N and 'Bart's'. Brit. Med. J. *1*, 929.

AGER, J. A. M., LEHMANN, H. and VELLA, F. (1958). Haemoglobin 'Norfolk': a new haemoglobin found in an English family. Brit. Med. J. *2*, 529.

ALLAN, J. D., CUSWORTH, D. L., DENT, C. E. and WILSON, V. D. (1958). A disease, probably hereditary, characterised by severe mental deficiency and a constant gross abnormality of aminoacid metabolism. Lancet *1*, 182.

ALLISON, A. C. (1954). Protection afforded by the sickle cell trait against subtertian malarial infection. Brit. Med. J. *1*, 290.

ALLISON, A. C. (1964). Polymorphism and natural selection in human populations. Cold Spring Harbor Symp. Quant. Biol. *29*, 137.

ALLISON, A. C. and BLUMBERG, B. (1961). An isoprecipitation reaction distinguishing human serum protein types. Lancet *1*, 634.

ALPER, C. A. and PROPP, R. P. (1968). Genetic polymorphism of the third component of human complement ($C^{'}3$). J. Clin. Invest. *47*, 2181.

ANDERSEN, D. H. (1956). Familial cirrhosis of the liver with storage of abnormal glycogen. Lab. Invest. *5*, 11.

APPELLA, E. and MARKERT, C. L. (1961). Dissociation of lactate dehydrogenase into subunits with guanidine hydrochloride. Biochem. Biophys. Res. Commun. *6*, 171.

ARCHER, H. E., DORMER, A. E., SCOWEN, E. F. and WATTS, R. W. E. (1957). Primary hyperoxaluria. Lancet *2*, 320.

ARMSTRONG, M. D. and ROBINSON, K. S. (1954). On the excretion of indole derivatives in phenylketonuria. Arch. Biochem. *52*, 287.

ARMSTRONG, M. D., SHAW, K. N. F. and ROBINSON, K. S. (1955). Studies on phenylketonuria. II. The excretion of o-hydroxyphenyl-acetic acid in phenylketonuria. J. Biol. Chem. *213*, 797.

ARMSTRONG, M. D. and TYLER, F. H. (1955). Studies on phenylketonuria. I. Restricted phenylalanine intake in phenylketonuria. J. Clin. Invest. *34*, 565.

ARROW, V. K. and WESTALL, R. G. (1958). Aminoacid clearances in cystinuria. J. Physiol. (London) *142*, 141.

AUERBACH, V. H., DIGEORGE, A. M. and CARPENTER, G. C. (1967). Phenylalaninemia. *In:* Aminoacid metabolism and genetic variation, ed. W. L. Nyhan. McGraw Hill, New York.

AUERBACH, V. H., DIGEORGE, A. M., BALDRIDGE, R. C., TOURTELLOTTE, C. D. and BRIGHAM, M. P. (1962). Histidinemia. A deficiency in histidase resulting in the urinary excretion of histidine and of imidazolepyruvic acid. J. Pediat. *60*, 487.

AUSTIN, J., BALASUBRAMANIAN, A., PATTABIRAMAN, T., SARSWATHI, S., BASU, D. and BACHHAWAT, B. (1963). A controlled study of enzymic activities in three human disorders of glycolipid metabolism. J. Neurochem. *10*, 805.

AZEN, E. A. and SMITHIES, O. (1968). Genetic polymorphism of *C'*3 (β1C-globulin) in human serum. Science 162, 905.

AZEVEDO, E., KIRKMAN, H. N., MORROW, A. C. and MOTULSKY, A. G. (1968). Variants of red cell glucose-6-phosphate dehydrogenase among Asiatic Indians. Ann. Hum. Genet. Lond. *31*, 375.

BAEHNER, R. L. and KARNOVSKY, M. L. (1968). Deficiency of reduced nicotinamide-adenine dinucleotide oxidase in chronic granulomatous disease. Science *162*, 1277.

BAGLIONI, C. (1962a). The fusion of two polypeptide chains in haemoglobin Lepore and its interpretation as a genetic deletion. Proc. Natl. Acad. Sci. U.S. *48*, 1880.

BAGLIONI, C. (1962b). A chemical study of hemoglobin Norfolk. J. Biol. Chem. *237*, 69.

BAGLIONI, C. (1963). A child homozygous for persistence of foetal haemoglobin. Nature (Lond.) *198*, 1117.

BAGLIONI, C. and INGRAM, V. M. (1961). Four adult haemoglobin types in one person. Nature (Lond.) *189*, 465.

BANK, A. and MARKS, A. (1966). Excess α chain synthesis relative to β chain synthesis in thalassaemia major and minor. Nature (Lond.) *212*, 1198.

BARCROFT, H., GIBSON, Q., HARRISON, D. and MCMURRAY, J. (1945). Familial idiopathic methaemoglobinaemia and its treatment with ascorbic acid. Clin. Sci. *5*, 145.

BARGELLESI, A., PONTREMOLI, S. and CONCONI, F. (1967). Absence of β globin synthesis and excess of α globin synthesis in homozygous β-thalassemia. Europ. J. Biochem. *1*, 73.

BARNABAS, J. and MULLER, C. J. (1962). Haemoglobin Lepore (Hollandia). Nature (Lond.) *194*, 931.

BARNES, H. D. (1958). Porphyria in South Africa: the faecal excretion of porphyrin. S. African Med. J. *32*, 680.

BARON, D. N., DENT, C. E., HARRIS, H., HART, E. W. and JEPSON, J. B. (1956). Hereditary pellagra-like skin rash with temporary cerebellar ataxia, constant renal aminoaciduria, and other bizarre biochemical features. Lancet *2*, 421.

BAUDHUIN, P., HERS, H. G. and LOEB, H. (1964). An electron microscopic and biochemical study of type II glycogenosis. Lab. Invest. *13*, 1140.

BAUGHAN, M. A., VALENTINE, W. N., PAGLIA, M. D., WAYS, P. O., SIMON, E. R. and DE MARSH Q. B. (1968). Hereditary hemolytic anemia associated with glucosephosphate isomerase (GPI) deficiency – a new enzyme defect of human erythrocytes. Blood *32*, 236.

BEARN, A. F., BOWMAN, B. H. and KITCHIN, F. D. (1964). Genetical and biochemical considerations of the serum group-specific component. Cold Spring Harbor Symp. Quant. Biol. *29*, 435.

BEET, E. A. (1949). The genetics of sickle cell trait in a Bantu tribe. Ann. Eugenics *14*, 279.

BELL, J. C. and RIEMENSNIDER, D. K. (1957). Use of a serum microbiologic assay technique for estimating patterns of isoniazid metabolism. Ann. Rev. Tuberc. *75*, 995.

BENESCH, R., BENESCH, R. E. and TYUMA, I. (1966). Subunit exchange and ligand binding. II. The mechanism of the allosteric effect in hemoglobin. Proc. Natl. Acad. Sci. U.S. *56*, 1268.

BERETTA, A., PRATO, V., GALLO, E. and LEHMANN, H. (1968). Haemoglobin Torino – α43 (CD1) phenylalanine → valine. Nature (Lond.) *217*, 1016.

BERG, K. (1963). A new serum type system in man – the Lp system. Acta Pathol. Microbiol. Scand. *59*, 369.

BERG, K. (1965). A new serum type system in man – the Ld system. Vox Sanguinis *10*, 513.

BERG, K. (1966). Further studies on the Lp system. Vox Sanguinis *11*, 419.

BERG, K. (1968). The Lp system. Series Haematologica *1*, 111.

BERG, K. and BEARN, A. G. (1966). An inherited X-linked serum system in man. The Xm System. J. Expl. Med. *123*, 379.

BERG, K. and BEARN, A. G. (1968). Human serum protein polymorphisms. Ann. Rev. Genet. *2*, 341.

BERMAN, J. L., CUNNINGHAM, G. C., DAY, R. W., FORD, R. and HSIA, D. Y. Y. (1969). Causes for high phenylalanine with normal tyrosine. Am. J. Diseases Children *117*, 54.

BERRY, H., SUTHERLAND, B. S. and GUEST, G. M. (1957). Phenylalanine tolerance tests on relatives of phenylketonuric children. Am. J. Human Genet. *9*, 310.

BEUTLER, E. and BALUDA, M. C. (1964). The separation of glucose-6-phosphate dehydrogenase deficient erythrocytes from the blood of heterozygotes for glucose-6-phosphate dehydrogenase deficiency. Lancet *1*, 189.

BEUTLER, E., BALUDA, M. C., STURGEON, P. and DAY, R. (1965). A new genetic abnormality resulting in galactose-1-phosphate uridyltransferase deficiency. Lancet *1*, 353.

BEUTLER, E., BALUDA, M. C., STURGEON, P. and DAY, R. (1966). The genetics of galactose-1-phosphate uridyl transferase deficiency. J. Lab. Clin. Med. *64*, 646.

BEUTLER, E., MATHAI, C. K. and SMITH, J. E. (1968). Biochemical variants of glucose-6-phosphate dehydrogenase giving rise to congenital non spherocytic hemolytic disease. Blood *31*, 131.

BEUTLER, E., YEH, M. and FAIRBANKS, V. F. (1962). The normal human female as a mosaic of X chromosome activity: studies using the gene for G-6-PD deficiency as a marker. Proc. Natl. Acad. Sci. U.S. *48*, 9.

BHENDE, Y., DESPHANDE, C. K., BHATIA, H. M., SANGER, R., RACE, R. R., MORGAN, W. T. J. and WATKINS, W. M. (1952). A 'new' blood group character related to the ABO system. Lancet *1*, 903.

BIANCO, A. and ZINKHAM, W. H. (1963). Lactate dehydrogenases in human testes. Science *139*, 601.

BIANCO, I., MONTALENTI, G., SILVESTRONI, E. and SINISCALCO, M. (1952). Further data on genetics of microcythaemia or thalassaemia minor and Cooley's disease or thalassaemia major. Ann. Eugenics (Lond.) *16*, 299.

BIANCO, A., ZINKHAM, W. H. and KUPCHYK, L. (1964). Genetic control and ontogeny of lactate dehydrogenase in pigeon testes. J. Exptl. Zool. *156*, 137.

BIGLEY, R. H. and KOLER, R. D. (1968). Liver pyruvate kinase (PK) isozymes in a PK-deficient patient. Ann. Hum. Genet. Lond. *31*, 383.

BIGLEY, R. H., STENZEL, P., JONES, R. T., CAMPOS, J. O. and KOLER, R. D. (1968). Tissue distribution of human pyruvate kinase isozymes. Enzymol. Biol. Clin. *9*, 10.

BLACK, J. A. and DIXON, G. H. (1968). Aminoacid sequence of alpha chains of human haptoglobins. Nature (Lond.) *218*, 736.

BLACK, J. A. and SIMPSON, K. (1967). Fructose intolerance. Brit. Med. J. *4*, 138.

BLUMBERG, B. S., BERNANKE, D. and ALLISON, A. C. (1962). A human lipoprotein polymorphism. J. Clin. Invest. *41*, 1936.

BOIVIN, P. and GALAND, C. (1967). Constant de Michaelis anormale pour le phospho-énol-pyruvate au cours d'un déficit en pyruvate-kinase erythrocytaire. Rev. Franc. Etudes Clin. Biol. *12*, 372.

BOOKCHIN, R. M., NAGEL, R. L. and RANNEY, H. M. (1967). Structure and properties of Hb C Harlem, a human haemoglobin variant with aminoacid substitutions in two residues of the β polypeptide chain. J. Biol. Chem. *242*, 248.

BOURNE, J. G., COLLIER, H. O. J. and SOMERS, G. F. (1952). Succinylcholine (succinoylcholine). Muscle relaxant of short action. Lancet *1*, 1225.

BOWMAN, E., FRISCHER, H., AJMAR, F., CARSON, P. and GOWER, M. K. (1967). Population, family and biochemical investigation of human adenylate kinase polymorphism. Nature (Lond.) *214*, 1156.

BOWMAN, H. S. and PROCOPIO, F. (1963). Hereditary non-spherocytic haemolytic anaemia of the pyruvate kinase deficient type. Ann. Internal Med. *58*, 567.

BOYER, S. H. (1961). Alkaline phosphatase in human sera and placentae. Science *134*, 1002.

BOYER, S. H., FAINER, D. C. and WATSON-WILLIAMS, E. J. 1963 Lactate dehydrogenase variant from human blood: evidence for molecular subunits. Science *141*, 642.

BOYER, S. H., HATHAWAY, P. and GARRICK, M. D. (1964). Modulation of protein synthesis in man: an *in vitro* study of haemoglobin synthesis by heterozygotes. Cold Spring Harbor Symp. Quant. Biol. *29*, 333.

BOYER, S. H., PORTER, I. H. and WEILBACHER, R. G. (1962). Electrophoretic heterogeneity of glucose-6-phosphate dehydrogenase and its relationship to enzyme deficiency in man. Proc. Natl. Acad. Sci. U.S. *48*, 1868.

BRADLEY, T. B., BOYER, S. H. and ALLEN, F. H. (1961). Hopkins-2-hemoglobin: a revised pedigree with data on blood and serum groups. Bull. Johns Hopkins Hosp. *108*, 75.

BRADLEY, T. B. JR., BRAWNER, J. N. III and CONLEY, C. L. (1961). Further observations on an inherited anomaly characterised by persistance of fetal hemoglobin. Bull. Johns Hopkins Hosp. *108*, 242.

BRADLEY, T. B. JR., WOHL, R. C. and RIEDER, R. F. (1967). Hemoglobin Gun Hill: deletion of five amino acid residues and impaired heme-globin binding. Science *157*, 1581.

BRADY, R. O. (1968). Enzymatic defects in the sphingolipidoses. Advan. Clin. Chem. *11*, 1.

BRADY, R. O., GAL, A. E., BRADLEY, R. M. and MARTENSSON, E. (1967a). The metabolism of ceramidetrihexosides. I. Purification and properties of an enzyme that cleaves the terminal galactose molecule of galactosylgalactosylglucosylceramide. J. Biol. Chem. *242*, 1021.

BRADY, R. O., GAL, A. E., BRADLEY, R. M., MARTENSSON, E., WARSHAW, A. L. and LASTER, L.

(1967b). Enzymatic defect in Fabry's disease. Ceramidetrihexosidase deficiency. New Engl. J. Med. *276*, 1163.

BRADY, R. O., KANFER, J. N., MOCK, M. B. and FREDRICKSON, D. S. (1966a). The metabolism of sphingomyelin. II. Evidence of an enzymatic deficiency in Niemann-Pick disease. Proc. Natl. Acad. Sci. U.S. *55*, 366.

BRADY, R. O., KANFER, J. N. and SHAPIRO, D. (1965). The metabolism of glycocerebrosides. II. Evidence of an enzymatic deficiency in Gaucher's disease. Biochem. Biophys. Res. Commun. *18*, 221.

BRADY, R. O., KANFER, J. N., SHAPIRO, D. AND BRADLEY, R. M. (1966b). Demonstration of a deficiency of glucocerebroside-cleaving enzyme in Gaucher's disease. J. Clin. Invest. *45*, 1112.

BRANCATI, C. and BAGLIONI, C. (1966). Homozygous $\beta\delta$ thalassaemia ($\beta\delta$-microcythaemia). Nature (Lond.) *212*, 262.

BRAUNITZER, G., HILSE, K., RUDOLFF, V. and HILSCHMANN, N. (1964). The haemoglobins. Advan. Protein Chem. *19*, 1.

BRAVERMAN, A. S. and BANK, A. (1969). Changing rates of globin chain synthesis during erythroid cell maturation in thalassaemia. J. Mol. Biol. *42*, 57.

BREMER, H. J. and NEUMANN, W. (1966). Tolerance of phenylalanine after intravenous administration in phenylketonurics, heterozygous carriers, and normal adults. Nature (Lond.) *209*, 1148.

BRENTON, D. P., CUSWORTH, D. C., DENT, C. E. and JONES, E. E. (1966). Homocystinuria. Clinical and dietary studies. Quart. J. Med. XXXV, 325.

BREWER, G. J. (1967). Achromatic regions of tetrazolium stained starch gels: inherited electrophoretic variation. Am. J. Human Genet. *19*, 674.

BREWER, G. J., EATON, J. W., KNUTSEN, C. S. and BECK, C. C. (1967). A starch-gel electrophoretic method for the study of diaphorase isozymes and preliminary results with sheep and human erythrocytes. Biochem. Biophys. Res. Commun. *29*, 198.

BROWN, B. I. and BROWN, D. H. (1966). Lack of an α-1, 4-glucan: α-1, 4-glucan 6-glycosyl transferase in a case of Type IV glycogenosis. Proc. Natl. Acad. Sci. U.S. *56*, 725.

BUDD, M. A., TANAKA, K., HOLMES, L. B., EFRON, M. L., CRAWFORD, J. D. and ISSELBACHER, K. J. (1967). Isovaleric acidemia. Clinical features of a new genetic defect of leucine metabolism. New Engl. J. Med. *277*, 321.

BUNDSCHUH, G. and VOGT, A. (1965). Die Häufigkeit des Merkmals Lp (x) in der Berliner Bevölkerung. Humangenetik *1*, 379.

BURNS, J. J. (1957). Missing step in man, monkey and guinea-pig required for biosynthesis of L-ascorbic acid. Nature (Lond.) *180*, 553.

BÜTLER, R. (1967). Polymorphisms of the human low-density lipoproteins. Vox Sanguinis *12*, 2.

CAHN, R. D., KAPLAN, N. O., LEVINE, L. and ZWILLING, E. (1962). Nature and development of lactic dehydrogenases. Science *136*, 962.

CARRELL, R. W., LEHMANN, H. and HUTCHISON, H. E. (1966). Haemoglobin Köln (β-98 Valine $\rightarrow$ Methionine): an unstable protein causing inclusion-body anaemia. Nature (Lond.) *210*, 915.

CARSON, N. A. J., DENT, C. E., FIELD, C. M. B. and GAULL, G. E. (1965). Homocystinuria. Clinical and pathological review of ten cases. J. Pediat. *66*, 565.

CARSON, P. E., FLANAGAN, C. L., ICKES, C. E. and ALVING, A. S. (1956). Enzymatic deficiency in primaquine sensitive erythrocytes. Science *124*, 484.

CARTER, N. D., FILDES, R. A., FITCH, L. I. and PARR, C. W. (1968). Genetically determined electrophoretic variations of human phosphogluconate dehydrogenase. Acta Genet. *18*, 109.

CEPPELLINI, R. (1955). Nuova interpretazione sulla genetica dei carratteri Lewis eritrocitari e salivari derivante dall'analysi di 87 famiglie. Ric. Sci. Suppl. *25*, 3.

CEPPELLINI, R. (1959). *In:* Biochemistry of human genetics. Ciba Foundation Symposium, ed. G. E. W. Wolstenholme and C. M. O'Connor. p. 133. Churchill, London.

CHAMBERS, R. A. and PRATT, R. T. C. (1956). Idiosyncrasy to fructose. Lancet *2*, 340.

CHILDS, B., ZINKHAM, W., BROWN, E. A., KIMBRO, E. L. and TORBET, J. V. (1958). A genetic study of a defect in glutathione metabolism of the erythrocyte. Bull. Johns Hopkins Hosp. *102*, 21.

CLARK, S. W., GLAUBIGER, G. A. and LA DU, B. N. (1968). Properties of plasma cholinesterase variants. Ann. N.Y. Acad. Sci. *151*, 710.

CLEGG, J. B., WEATHERALL, D. J., NA-NAKORN, S. and WASI, P. (1968). Haemoglobin synthesis in β thalassaemia. Nature (Lond.) *220*, 664.

CLEVE, H., GORDON, S., BOWMAN, B. H. and BEARN, A. G. (1967). Comparison of the tryptic peptides and amino acid composition of the beta polypeptide chains of the three common haptoglobin phenotypes. Am. J. Human Genet. *19*, 713.

COMINGS, D. E. and MOTULSKY, A. G. (1966). Absence of cis delta chain synthesis in ($\delta\beta$) thalassaemia (F-thalassaemia). Blood *28*, 54.

CONLEY, C. L., WEATHERALL, D. J., RICHARDSON, S. N., SHEPHERD, M. K. and CHARACHE, C. (1963). Hereditary persistence of fetal hemoglobin: a study of 79 affected persons in 15 Negro families in Baltimore. Blood *21*, 261.

CONNELL, G. E., DIXON, G. H. and SMITHIES, O. (1962). Subdivision of the three common haptoglobin types based on hidden differences. Nature (Lond.) *193*, 505.

CONNELL, G. E., SMITHIES, O. and DIXON, G. H. (1966). Gene action in the human haptoglobins. II. Isolation and physical characterisation of alpha polypeptide chains. J. Mol. Biol. *21*, 225.

CORI, G. T. (1954). Glycogen structure and enzyme deficiencies in glycogen storage diseases. Harvey Lectures Ser. *48*, p. 145.

CORI, G. T. (1957). Biochemical aspects of glycogen deposition diseases. Mod. Probl. Pediat. *3*, 344.

CORI, G. T. and CORI, C. F. (1952). Glucose-6-phosphatase of the liver in glycogen storage disease. J. Biol. Chem. *199*, 661.

COTE, R. H. and MORGAN, W. T. J. (1956). Some nitrogen-containing disaccharides isolated from human blood-group A substances. Nature (London) *178*, 1171.

COURT-BROWN, W. M. and SMITH, P. G. (1969). Human population cytogenetics. Brit. Med. Bull *25*, 74.

CRICK, F. H. C. (1967). The genetic code. Proc. Roy. Soc. B *167*, 331.

CURRARINO, G., NEWHAUSER, E., REYERSBACK, G. and SOBEL, E. (1957). Hypophosphatasia. Am. J. Roentgenol. *78*, 392.

CURTAIN, C. C. (1964). A structural study of abnormal haemoglobins occurring in New Guinea. Australian J. Exptl. Biol. Med. Sci. *42*, 89.

DACIE, J. V., SHINTON, N. K., GAFFNEY, P. J., CARRELL, R. W. and LEHMANN, H. (1967). Haemoglobin Hammersmith (β42 (CD1) Phe → Ser). Nature (Lond.) *216*, 663.

DACREMONT, G. and KINT, J. A. (1968). Gm_1 – ganglioside accumulation and β-galactosidase deficiency in a case of Gm_1 gangliosidosis (Landing disease). Clin. Chim. Acta *21*, 421.

DANCIS, J., HUTZLER, J. and LEVITZ, M. (1963). The diagnosis of maple syrup urine disease (branch chain ketoaciduria) by the *in vitro* study of the peripheral leucocyte. Pediatrics *32*, 234.

DANCIS, J., LEVITZ, M., MILLER, S. and WESTALL, R. G. (1959). Maple syrup urine disease. Brit. Med. J. *1*, 91.

DANCIS, J., HUTZLER, J. and LEVITZ, M. (1965). Detection of the heterozygote in maple syrup urine disease. J. Pediat. *66*, 595.

DANCIS, J., HUTZLER, J., TADA, K., WADA, Y., MORIKAWA, T. and ARAKAWA, T. (1967). Hypervalinemia. A defect in valine transamination. Pediatrics *39*, 813.

DANCIS, J., HUTZLER, J. and ROKKONES, T. (1967). Intermittent branched-chain ketonuria. New Engl. J. Med. *276*, 84.

DANOVITCH, S. H., BAER, P. N. and LASTER, L. (1968). Intestinal alkaline phosphatase activity in familial hypophosphatasia. New Engl. J. Med. *278*, 1253.

DAVIDSON, R. G. and CORTNER, J. A. (1967a). Genetic variant of human erythrocyte malate dehydrogenase. Nature (Lond.) *215*, 761.

DAVIDSON, R. G. and CORTNER, J. A. (1967b). Mitochondrial malate dehydrogenase: a new genetic polymorphism in man. Science *157*, 1569.

DAVIDSON, R. G., FILDES, R. A., GLEN-BOTT, A. M., HARRIS, H., ROBSON, E. B. and CLEGHORN, T. E. (1965). Genetical studies on a variant of human lactate dehydrogenase (subunit A). Ann. Hum. Genet. Lond. *29*, 5.

DAVIDSON, R. G., NITOWSKY, J. M. and CHILDS, B. (1963). Demonstration of two populations of cells in the human female heterozygous for glucose-6-phosphate dehydrogenase variants. Proc. Natl. Acad. Sci. U.S. *50*, 481.

DAVIES, R. O., MARTON, A. V. and KALOW, W. (1960). The action of normal and atypical cholinesterase of human serum upon a series of esters of choline. Can. J. Biochem. Physiol. *38*, 545.

DEAN, G. (1963). The porphyrias: a story of inheritance and environment. Pitman, London.

DENT, C. E. and PHILPOT, G. R. (1954). Xanthinuria, an inborn error (or deviation) of metabolism. Lancet *1*, 182.

DENT, C. E. and ROSE, G. A. (1951). Aminoacid metabolism in cystinuria. Quart. J. Med. *20*, 205.

DENT, C. E. and SENIOR, B. (1955). Studies on the treatment of cystinuria. Brit. J. Urol. *27*, 317.

DENT, C. E., SENIOR, B. and WALSHE, J. M. (1954). The pathogenesis of cystinuria. II. Polarographic studies of the metabolism of sulphur-containing aminoacids. J. Clin. Invest. *33*, 1216.

DERN, R. J., MCCURDY, P. R. and YOSHIDA, A. (1969). A new structural variant of glucose-6-phosphate dehydrogenase with a high production rate (G6PD Hektoen). J. Lab. Clin. Med. *73*, 283.

DERN, R. J., WEINSTEIN, I. M., LEROY, G. V., TALMAGE, D. W. and ALVING, A. S. (1954). The hemolytic effect of primaquine. I. The localization of the drug-induced hemolytic defect in primaquine-sensitive individuals. J. Lab. Clin. Med. *43*, 303.

DETTER, J. C., WAYS, P. O., GIBLETT, E. R., BAUGHAN, M. A., HOPKINSON, D. A., POVEY, S. and HARRIS, H. (1968). Inherited variations in human phosphohexose isomerase. Ann. Hum. Genet. Lond. *31*, 329.

DEVADATTA, S., GANGADHARAM, P. R., ANDREWS, R. H., FOX, W., REMAKRISHNAN, C. V., SELKON, J. B. and VELU, S. (1960). Peripheral neuritis due to isoniazid. Bull. World Health Org. *23*, 587.

DIXON, G. H. (1966). Mechanisms of protein evolution. Essays in Biochemistry *2*, 147. Academic Press, New York.

DOENICKE, A., GARTNER, T., KRENZBERG, G., REMES, I., SPIESS, W. and STEINBEREITHNER, K. (1963). Serum cholinesterase anenzymia. Acta Anaesthesiol. Scand. *7*, 59.

DONNELL, G. N., BERGREN, W. R., BRETTHAUER, R. K. and HANSEN, R. G. (1960). The enzymatic expression of heterozygosity in families of children with galactosemia. Pediatrics *25*, 572.

DOOLAN, P. D., HARPER, H. A., HUTCHIN, M. E. and ALPEN, E. L. (1957). Renal clearance of lysine in cystinuria. Am. J. Med. *23*, 416.

DOWDLE, E. B., MUSTARD, P. and EALES, L. (1967). δ-aminolaevulinic acid synthetase activity in normal and porphyric human livers. S. African Med. J. *41*, 1093.

DUMA, H., EFREMOV, G., SADIKARIO, A., TEODOSIJEV, D., MLADENOVSKI, B., VLASKI, R. and ANDREEVA, M. (1968). Study of nine families with haemoglobin-Lepore. Brit. J. Haematol. *15*, 161.

DE DUVE, C. (1963). The lysosome concept. *In:* CIBA Foundation Symposium on Lysosomes. Churchill, London.

EDINGTON, G. M. and LEHMANN, H. (1954). Haemoglobin G: a new haemoglobin found in a West African. Lancet *2*, 173.

EDINGTON, G. M. and LEHMANN, H. (1955). Expression of the sickle cell gene in Africa. Brit. Med. J. *1*, 1308, and *2*, 1328.

EDINGTON, G. M. and WATSON-WILLIAMS, E. J. (1964). Sickling, haemoglobin C, glucose-6-phosphate dehydrogenase deficiency and malaria in Western Nigeria. *In:* Abnormal haemoglobins in Africa, ed. J. H. P. Jonxis. Blackwell, Oxford.

EFRON, M. L. (1965). Familial hyperprolinemia. Report of a second case, associated with congenital renal malformations, hereditary hematuria and mild mental retardation, with demonstration of an enzyme defect. New Engl. J. Med. *272*, 1243.

EFRON, M. L. (1965). Hydroxyprolinemia. II. A rare metabolic disease due to a deficiency of the enzyme 'hydroxyproline oxidase'. New Engl. J. Med. *272*, 1299.

EFRON, M. L. (1966). Diseases of the urea cycle. *In:* The metabolic basis of inherited disease, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. McGraw-Hill, New York.

ENGLEMAN, K., WATTS, R. W. E., KLINENBERG, J. R., SJOERDSMA, A. and SEEGMILLER, J. E. (1964). Clinical, physiological and biochemical studies of a patient with xanthinuria and pheochromocytoma. Am. J. Med. *37*, 839.

ENCKLEWITZ, M. and LASKER, M. (1935). The origin of L-xyloketose (urine pentose). J. biol. Chem. *110*, 443.

EPSTEIN, C. J. and SCHECHTER, A. N. (1968). An approach to the problem of conformational isozymes. Ann. New York Acad. Sci. *151*, 85.

EVANS, D. A. P. (1963). Pharmacogenetics. Am. J. Med. *34*, 639.

EVANS, D. A. P. and CLARKE, C. A. (1961). Pharmacogenetics. Brit. Med. Bull. *17*, 234.

EVANS, D. A. P., DAVIDSON, K. and PRATT, R. T. C. (1965). The influence of acetylator phenotype on the effect of treating depression with phenelzine. Clin. Pharmacol. Therap. *6*, 430.

EVANS, D. A. P., MANLEY, K. E. and MCKUSICK, V. A. (1960). Genetic control of isoniazid metabolism in man. Brit. Med. J. *2*, 485.

EVANS, D. A. P., STOREY, P. B. and MCKUSICK, V. A. (1961). Further observations on the determination of the isoniazid inactivator phenotype. Bull. Johns Hopkins Hosp. *108*, 60.

EVANS, D. A. P. and WHITE, T. A. (1964). Human acetylation polymorphism. J. Lab. Clin. Med. *63*, 394.

EVANS, F. T., GRAY, P. W. S., LEHMANN, H. and SILK, E. (1952). Sensitivity to succinyl-choline in relation to serum cholinesterase. Lancet *1*, 1229.

FAGERHOL, M. K. (1967). Serum Pi types in Norwegians. Acta Pathol. Microbiol. Scand. *70*, 421.

FAGERHOL, M. K. (1968). The P_1 system: genetic variants of serum α_1 antitrypsin. Series Haematologica *1*, 153.

FAGERHOL, M. K. (1969). Quantitative studies on the inherited variants of serum α-antitrypsin. Scand. J. Clin. Lab. Invest. *23*, 97.

FAGERHOL, M. K. and BRAEND, M. (1965). Serum prealbumin: polymorphism in man. Science 149, 986.

FAGERHOL, M. K. and LAURELL, C. B. (1967). The polymorphism of 'prealbumins' and α_1-antitrypsin in human sera. Clin. Chim. Acta *16*, 199.

FALLON, H. J., SMITH, L. H., GRAHAM, J. B. and BURNETT, C. H. (1964). A genetic study of hereditary orotic aciduria. New Engl. J. Med. *270*, 878.

FESSAS, P. and STAMATOYANNOPOULOS, G. (1964). Hereditary persistence of foetal haemoglobin in Greece. A study and comparison. Blood *24*, 223.

FILDES, R. A. and HARRIS, H. (1966). Genetically determined variation of adenylate kinase in man. Nature (Lond.) *209*, 261.

FILDES, R. A. and PARR, C. W. (1963). Human red cell phosphogluconate dehydrogenase. Nature (Lond.) *200*, 890.

FINKELSTEIN, J. D., MUDD, S. H., IRREVERRE, F. and LASTER, L. (1966). Deficiencies of cystathionase and homoserine dehydratase activities in cystathioninuria. Proc. Natl. Acad. Sci. U.S. *55*, 865.

FISHER, R. A. (1930). The genetical theory of natural selection. Clarendon Press, Oxford.

FISHER, R. A. and HARRIS, H. (1969). Studies on the purification and properties of the genetic variants of red cell acid phosphohydrolase in man. Ann. New York Acad. Sci. *166* 380.

FÖLLING, A. (1934). Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechselanomalie in Verbindung mit Imbezillität. Hoppe-Seylers Z. Physiol. Chem. *227*, 169.

FORBES, G. B. (1953). Glycogen disease. Report of a case with abnormal glycogen structure in liver and skeletal muscle. J. Pediat. *42*, 645.

FRASER, D. (1957). Hypophosphatasia. Am. J. Med. *22*, 730.

FREDRICKSON, D. S. (1966). Sphingomyelin lipidosis: Niemann–Pick disease. *In:* The metabolic basis of inherited disease, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. 2nd ed. McGraw-Hill, New York.

FRIMPTER, G. W. (1965). Cystathioninuria: nature of the defect. Science *149*, 1095.

FRIMPTER, G. W. (1967). Cystathioninuria. *In:* Aminoacid metabolism and genetic variation, ed. W. L. Nyhan. McGraw-Hill, New York.

FRIMPTER, G. W., HAYMOVITZ, A. and HORWITH, M. (1963). Cystathioninuria. New Engl. J. Med. *268*, 333.

FROESCH, E. R., PRADER, A., LABHART, A., STUBER, H. W. and WOLF, H. P. (1957). Die hereditäre Fructoseinteroleranz, einer bisher nicht bekannte kongenitaler Stoffwechselstörung. Schweiz. Med. Wschr. *87*, 1168.

FROESCH, E. R., WOLF, H. P., BAITSCH, H., PRADER, A. and LABHART, A. (1963). Hereditary fructose intolerance. An inborn defect of hepatic fructose-1-phosphate splitting aldolase. Am. J. Med. *34*, 151.

GARROD, A. E. (1909). Inborn errors of metabolism. Oxford University Press.

GENTZ, J., JAGENBURG, R. and ZETTERSTRÖM, R. (1965). Tyrosinemia: an inborn error of tyrosine metabolism with cirrhosis of the liver and multiple renal tubular defects. J. Pediat. *66*, 620.

GERALD, P. S. and DIAMOND, L. K. (1958). A new hereditary haemoglobinopathy (the Lepore trait) and its interaction with the thalassaemia trait. Blood *12*, 835.

GERALD, P. S. and EFRON, M. L. (1961). Chemical studies of several varieties of Hb-M. Proc. Natl. Acad. Sci. U.S. *47*, 1758.

GERALD, P. S. and SCOTT, E. M. (1966). The hereditary methaemoglobinaemias. *In:* The metabolic basis of hereditary disease, ed. J. B. Stanbury, J. B. Wyngaarden and S. A. Fredrickson. 2nd ed. McGraw-Hill, New York.

GERRITSEN, T. and WAISMAN, H. A. (1964a). Homocystinuria: absence of cystathionine in the brain. Science *145*, 588.

GERRITSEN, T. and WAISMAN, H. A. (1964b). Homocystinuria: an error in the metabolism of methionine. Pediatrics *33*, 413.

GHANDIMI, H. and PARTINGTON, M. W. (1967). Salient features of histidinemia. Am. J. Diseases Children *113*, 83.

GHADIMI, H., PARTINGTON, M. W. and HUNTER, A. (1961). A familial disturbance of histidine metabolism. New Engl. J. Med. *265*, 221.

GIBLETT, E. R. (1962). The plasma transferrins. *In:* Progress in medical genetics, Vol. 2, ed. A. G. Bearn and A. G. Steinberg. Grune and Stratton, New York.

GIBLETT, E. R. (1969). Genetic markers in human blood. Blackwell, Oxford.

GIBLETT, E. R. and SCOTT, N. M. (1965). Red cell acid phosphatase: racial distribution and report of a new phenotype. Am. J. Hum. Genet. *17*, 425.

GIBLETT, E. R. and STEINBERG, A. G. (1960). The inheritance of serum haptoglobin types of American Negroes: evidence for a third allele Hp^{2M}. Am. J. Hum. Genet. *12*, 160.

GIBSON, Q. (1948). The reduction of methaemoglobin in red blood cells and studies on the cause of idiopathic methaemoglobinaemia. Biochem. J. *42*, 13.

GILBERT, W. and MÜLLER-HILL, B. (1966). Isolation of the lac repressor. Proc. natl. Acad. Sci. U.S. *56*, 1891.

GILLES, H. M., FLETCHER, K. A., HENDRICKSE, R. G., LINDNER, R., REDDY, S. and ALLAN, N. (1967). Glucose-6-phosphate dehydrogenase deficiency, sickling, and malaria in African children in South Western Nigeria. Lancet *1*, 138.

GITZELMANN, R. (1967). Hereditary galactokinase deficiency, a newly recognised cause of juvenile cataracts. Pediat. Res. *1*, 14.

GITZELMANN, R., CURTIUS, H.-C., SCHNELLER, I. (1967). Galactitol and galactose-1-phosphate in the lens of a galactosemic infant. Exptl. Eye Res. *6*, 1.

GJONE, E. and NORUM, K. R. (1968). Familial serum cholesterol ester deficiency. Acta Med. Scand. *183*, 107.

GOEDDE, H. W., GEHRING, D. and HOFMANN, R. A. (1965). On the problem of a 'silent gene' in pseudocholinesterase polymorphism. Biochim. Biophys. Acta *107*, 391.

GOEDDE, H. W. and KELLER, W. (1967). Metabolic pathways in maple syrup urine disease. *In:* Amino acid metabolism and genetic variation, ed. W. L. Nyhan. McGraw-Hill, New York.

GOEDDE, H. W., RICHTER, E., HÜFNER, M. and VON ZUR MÜHLEN, A. (1964a). Untersuchungen zur Ahornsirupkrankheit an zwei Familien. Humangenetik *1*, 163.

GOEDDE, H. W., RICHTER, E., HÜFNER, M. and SIXEL, B. (1964b). Arbeitsvorschrift eines vereinfachten Heterozygotentestes für die Ahornsirup-Krankheit. Klin. Wschr. *15*, 818.

GOODMAN, S. I., MCINTYRE, C. A. and O'BRIEN, D. (1967). Impaired intestinal transport of proline in a patient with familial iminoaciduria. J. Pediat. *71*, 246.

GRANICK, J. (1966). The induction *in vitro* of the synthesis of δ-aminolaevulinic acid synthetase in chemical porphyria: a response to certain drugs, sex hormones and foreign chemicals. J. Biol. Chem. *241*, 1359.

GRIMES, A. J., MEISLER, A. and DACIE, J. V. (1964). Hereditary non-spherocytic haemolytic anaemia. A study of red cell carbohydrate metabolism in twelve cases of pyruvate kinase deficiency. Brit. J. Haematol. *10*, 403.

GRUBB, R. (1948). Correlation between Lewis blood group and secretor character in man. Nature (Lond.) *162*, 933.

GRUBB, R. (1951). Observations on the human group system Lewis. Acta Pathol. Microbiol. Scand. *28*, 61.

GRUBB, R. (1956). Agglutination of erythrocytes coated with 'incomplete' anti-Rh by certain rheumatoid arthritis sera and some other sera. The existence of human serum groups. Acta Pathol. Microbiol. Scand. *39*, 195.

GRUBB, R. and LAURELL, A. B. (1956). Hereditary serological human serum groups. Acta Pathol. Microbiol. Scand. *39*, 390.

GRUMBACH, M. M., MARKS, P. A. and MOROSHIMA, A. (1962). Erythrocyte glucose-6-phosphate dehydrogenase activity and X chromosome polysomy. Lancet *1*, 1330.

GUIDOTTI, G., KONIGSBERG, W. and CRAIG, L. C. (1963). On the dissociation of normal adult human haemoglobin. Proc. Natl. Acad. Sci. U.S. *50*, 774.

GUTSCHE, B. B., SCOTT, E. M. and WRIGHT, R. C. (1967). Hereditary deficiency of pseudocholinesterase in Eskimos. Nature (Lond.) *215*, 322.

HADORN, B., TARLOW, M. J., LLOYD, J. K. and WOLFF, O. H. (1969). Intestinal enterokinase deficiency. Lancet *1*, 812.

HAHN, E. V. and GILLESPIE, E. B. (1927). Sickle-cell anaemia: report of a case greatly improved by splenectomy; experimental study of sickle-cell formation. Arch. Int. Med. *39*, 233.

HAKOMORI, S. and STRYCHARZ, G. D. (1968). Investigations on cellular blood-group substances. I. Isolation and chemical composition of blood-group ABH and Le^b isoantigens of sphingoglycolipid nature. Biochem. *7*, 1279.

HALL CRAGGS, M., MARSDEN, P. D., RAPER, A. B., LEHMANN, H. and BEALE, D. (1964). Homozygous sickle cell anaemia arising from two different haemoglobins S. Brit. med. J. *2*, 87.

HALVORSEN, S., PANDE, H., LOKEN, A. C. and GJESSING, L. R. (1966). Tyrosinosis. Arch. Disease Childhood *41*, 238.

HAMNSTROM, B., GJONE, E. and NORUM, K. R. (1969). Familial plasma lecithin: cholesterol acyltransferase deficiency. Brit. Med. J. *2*, 283.

HARRAP, G. J. and WATKINS, W. M. (1964). Characterisation of the enzyme from *T. foetus* that destroys the serological specificity of blood-group A substance. Biochem. J. *93*, 9P.

HARRIS, H. (1966). Enzyme polymorphisms in man. Proc. Roy. Soc. B *164*, 298.

HARRIS, H. (1969). Genes and isozymes. Proc. Roy. Soc. B *174*, 1.

HARRIS, H., HOPKINSON, D. A. and ROBSON, E. B. (1962). Two dimensional electrophoresis of pseudo-cholinesterase components in normal human serum. Nature (Lond.) *196*, 1296.

HARRIS, H., HOPKINSON, D. A., ROBSON, E. B. and WHITTAKER, M. (1963a). Genetical studies on a new variant of serum cholinesterase detected by electrophoresis. Ann. Hum. Genet. Lond. *26*, 359.

HARRIS, H., HOPKINSON, D. A., LUFFMAN, J. E. and RAPLEY, S. (1968). Electrophoretic variation in red cell enzymes. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

HARRIS, H., HOPKINSON, D. A., SPENCER, N., COURT-BROWN, W. M. and MANTLE, D. (1963b). Red cell glucose-6-phosphate dehydrogenase activity in individuals with abnormal numbers of X-chromosomes. Ann. Hum. Genet. Lond. *27*, 59.

HARRIS, H., MITTWOCH, U., ROBSON, E. B. and WARREN, F. L. (1955a). Pattern of aminoacid excretion in cystinuria. Ann. Hum. Genet. Lond. *19*, 196.

HARRIS, H., MITTWOCH, U., ROBSON, E. B. and WARREN, F. L. (1955b). Phenotypes and genotypes in cystinuria. Ann. Hum. Genet. Lond. *20*, 57.

HARRIS, H., PENROSE, L. S. and THOMAS, D. H. H. (1959). Cystathioninuria. Ann. Hum. Genet. Lond. *23*, 442.

HARRIS, H. and ROBSON, E. B. (1959). A genetical study of ethanolamine phosphate excretion in hypophosphatasia. Ann. Hum. Genet. Lond. *23*, 421.

HARRIS, H. and ROBSON, E. B. (1963). Fractionation of human serum cholinesterase components by gel filtration. Biochim. Biophys. Acta *73*, 649.

HARRIS, H., ROBSON, E. B., GLEN-BOTT, A. M. and THORNTON, J. A. (1963c). Evidence for non-allelism between genes affecting human serum cholinesterase. Nature (Lond.) *200*, 1185.

HARRIS, H. and WARREN, F. L. (1953). Quantitative studies on the urinary cystine in patients with cystine stones and their relatives. Ann. Eugen. Lond. (now Ann. Hum. Genet.) *18*, 125.

HARRIS, H. and WHITTAKER, M. (1961). Differential inhibition of human serum cholinesterase with fluoride. Recognition of two new phenotypes. Nature (Lond.) *191*, 496.

HARRIS, H. and WHITTAKER, M. (1962). The serum cholinesterase variants: a study of twenty-two families selected via the 'intermediate' phenotype. Ann. Hum. Genet. Lond. *26*, 73.

HARRIS, H., WHITTAKER, M., LEHMANN, H. and SILK, E. (1960). The pseudocholinesterase

variants. Esterase levels and dibucaine numbers in families selected through suxamethonium sensitive individuals. Acta Genet. Stat. Med. *10*, 1.

HARRIS, J. W. (1950). Studies on the destruction of red blood cells. VIII. Molecular orientation in sickle cell haemoglobin solutions. Proc. Soc. Exp. Biol. N.Y. *75*, 197.

HARTMAN, G. (1941). Group antigens in human organs. Munksgaard, Copenhagen.

HARVALD, B., HANEL, K. H., SQUIRES, R. and TRAP-JENSEN, J. (1964). Adenosine-triphosphatase deficiency in patients with non-spherocytic haemolytic anaemia. Lancet *1*, 18.

HEARN, V. M., SMITH, Z. G. and WATKINS, W. M. (1968). An α-N-acetyl-D-galactosaminyltransferase associated with the human blood-group A character. Biochem. J. *109*, 315.

HERRICK, J. B. (1910). Peculiar elongated and sickle shaped red corpuscles in a case of severe anaemia. Arch. Intern. Med. *6*, 517.

HERS, H. G. (1959). Etudes enzymatiques sur fragments hépatiques. Application à la classification des glycogenoses. Rev. Intern. Hépatol. *9*, 35.

HERS, H. G. (1962). α-glucosidase deficiency in generalised glycogen-storage disease (Pompe's disease). Biochem. J. *86*, 11.

HERS, H. G. (1965). Inborn lysosomal diseases. Gastroenterology *48*, 625.

HERS, H. G. and JOASSIN, G. (1961). Anomalie de l'aldolase hépatique dans 'intolerance' au fructose. Enzymol. Biol. Clin. *1*, 4.

HIATT, H. H. (1966). Pentosuria. *In:* The metabolic basis of inherited disease, ed. J. S. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. 2nd edition. McGraw-Hill, New York.

HIRSCHFELD, J. (1959). Immunoelectrophoretic demonstration of qualitative differences in normal human sera and their relation to haptoglobins. Acta Pathol. microbiol. Scand. *47*, 160.

HIRSCHFELD, J. (1962). The Gc system. Immunoelectrophoretic studies of normal human sera with special reference to a new genetically determined serum system (Gc). Progr. Allergy *6*, 155.

HIRSCHFELD, J. (1968). The Ag system – a comparison of different isoprecipitation sera. Series Haematologica *1*, 38.

HIRSCHFELD, J. and OKACHI, K. (1967). Distribution of Ag(x) and Ag(y) antigens in some populations. Vox Sang. *13*, 1.

HOCKADAY, R. D. R., CLAYTON, J. E., FREDERICK, E. W. and SMITH, L. H. JR. (1964). Primary hyperoxaluria. Medicine *43*, 315.

HOCKWALD, R. S., ARNOLD, J., CLAYMAN, B. and ALVING, S. A. (1952). Toxicity of primaquine in Negroes. J. Am. Med. Ass. *149*, 1568.

HODGKIN, W. E., GIBLETT, E. R., LEVINE, H., BAUER, W. and MOTULSKY, A. G. (1965). Complete pseudocholinesterase deficiency: genetic and immunologic characterization. J. Clin. Invest. *44*, 486.

HOLLENDER, A., LORKIN, P. A., LEHMANN, H. and SVENSSON, B. (1969). New unstable haemoglobin Borås: β88 (F4) leucine → arginine. Nature (Lond.) *222*, 953.

HOLMES, B., QUIE, P. G., WINDHORST, D. B. and GOOD, R. A. (1966). Fatal granulomatous disease of childhood. An inborn error of phagocytic function. Lancet *1*, 1225.

HOLTON, J. B. (1965). Skin L-histidine ammonia-lyase activity in the family of a child with histidinaemia. Clin. Chim. Acta *11*, 193.

HOLZEL, A. and KOMROWER, G. M. (1955). A study of the genetics of galactosaemia. Arch. Disease Childhood *30*, 155.

HOLZEL, A., SCHWARTZ, V. and SUTCLIFFE, K. W. (1959). Defective lactose absorption causing malnutrition in infancy. Lancet 1, 1126.

HOOK, E. B., STAMATOYANNOPOULOS, G., YOSHIDA, A. and MOTULSKY, A. N. (1968). Glucose-6-phosphate dehydrogenase Madrona: a slow electrophoretic glucose-6-phosphate dehydrogenase variant with kinetic characteristics similar to those of normal type. J. Lab. Clin. Med. *72*, 404.

HOPKINSON, A. H. and HARRIS, H. (1965). Evidence for a second 'structural' locus determining human phosphoglucomutase. Nature (Lond.) *208*, 410.

HOPKINSON, D. A. and HARRIS, H. (1966). Rare phosphoglucomutase phenotypes. Ann. Hum. Genet. Lond. *30*, 167.

HOPKINSON, D. A. and HARRIS, H. (1968). A third phosphoglucomutase locus in man. Ann. Hum. Genet. Lond. *31*, 359.

HOPKINSON, D. A. and HARRIS, H. (1969). Red cell acid phosphatase, phosphoglucomutase and adenylate kinase. *In:* Biochemical methods in red cell genetics, ed. G. Yunis. Academic Press, New York.

HOPKINSON, D. A., SPENCER, N. and HARRIS, H. (1963). Red cell acid phosphatase variants; a new human polymorphism. Nature (Lond.) *199*, 969.

HOPKINSON, D. A., SPENCER, N. and HARRIS, H. (1964). Genetical studies on human red cell acid phosphatase. Am. J. Hum. Genet. *16*, 141.

HÖRLEIN, H. and WEBER, G. (1948). Über chronische familiäre Methämoglobinämia und eine neue Modifikation des Methämoglobins. Deutsch. Med. Wschr. *73*, 476.

HORTON, B. F. and HUISMAN, T. H. J. (1963). Linkage of the β-chain and γ-chain structural genes of human haemoglobins. Am. J. Hum. Genet. *15*, 394.

HOWELL, R., KLINENBERG, J. R. and KROOTH, R. S. (1967). Enzyme studies on diploid cell strains developed from patients with hereditary orotic aciduria. Johns Hopkins Med. J. *120*, 81.

HSIA, D. Y. Y., DRISCOLL, K., TROLL, W. and KNOX, W. E. (1956). Detection by phenylalanine tolerance tests of heterozygous carriers of phenylketonuria. Nature (Lond.) *178*, 1239.

HSIA, D. Y. Y., PAINE, R. S. and DRISCOLL, K. W. (1957). Phenylketonuria: Detection of the heterozygous carrier. J. Mental Deficiency Res. *1*, 53.

HUBBY, J. L. and LEWONTIN, R. C. (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. Genetics *54*, 577.

HUEHNS, E. R., DANCE, N., BEAVEN, G. H., HECHT, F. and MOTULSKY, G. (1964). Human embryonic haemoglobins. Cold Spring Harbor Symp. Quant. Biol. *29*, 327.

HUEHNS, E. R. and MODELL, C. B. (1967). Haemoglobin synthesis in thalassaemia. Trans. Roy. Soc. Trop. Med. Hyg. *61*, 157.

HUG, G., SCHUBERT, W. K. and CHUCK, G. (1966). Phosphorylase kinase of the liver: deficiency in a girl with increased hepatic glycogen. Science *153*, 1534.

HUGHES, H. P., BIEHL, J. P., JONES, A. P. and SCHMIDT, L. H. (1954). Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. Am. Rev. Tuberc. *70*, 266.

HUGH-JONES, K., NEWCOMB, A. L. and HSIA, D. Y. Y. (1960). The genetic mechanism of galactosaemia. Arch. Disease Childhood *35*, 521.

HUGULEY, C. M. JR., BAIN, J. A., RIVERS, S. and SCOGGINS, R. (1959). Refractory megaloblastic anaemia associated with excretion of orotic acid. Blood *14*, 615.

HUIJING, F., OBBINK, H. J. K. and VAN CREVELD, S. (1968). The activity of the debranching-enzyme system in leucocytes. A genetic study of glycogen storage disease type III. Acta Genet. Basel *18*, 128.

HUNT, J. A. and INGRAM, V. M. (1958). Allelomorphism and the chemical differences of the human haemoglobins A, S and C. Nature (Lond.) *184*, 1062.

HUNT, J. A. and LEHMANN, H. (1959). Haemoglobin Bart's: a foetal haemoglobin without α-chains. Nature (Lond.) *184*, 872.

HUNTER, R. L. and MARKERT, C. L. (1957). Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science *125*, 1294.

HUNTSMAN, R. G., HALL, M., LEHMANN, H. and SUKUMARAN, P. K. (1963). A second and a third abnormal haemoglobin in Norfolk. Brit. Med. J. *1*, 720.

HUTCHISON, J. H. and MCGIRR, E. M. (1956). Sporadic non-endemic goitrous cretinism. Lancet *1*, 1035.

ILLINGWORTH, B. and CORI, G. T. (1952). Structure of glycogens and amylopectins. III. Normal and abnormal human glycogen. J. Biol. Chem. *199*, 653.

ILLINGWORTH, B., CORI, G. T. and CORI, C. F. (1956). Amylo 1, 6 glucosidase activity in muscle tissue in generalised glycogen storage disease. J. Biol. Chem. *218*, 123.

INGRAM, V. M. (1957). Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. Nature (Lond.)*180*, 326.

INGRAM, V. M. (1959). Abnormal human haemoglobins. III. The chemical difference between normal and sickle cell haemoglobins. Biochim. Biophys. Acta *36*, 402.

INGRAM, V. M. (1961). Gene evolution and the haemoglobins. Nature (Lond.) *189*, 704.

INGRAM, V. M. and STRETTON, A. O. W. (1959). Genetic basis of the thalassemia diseases. Nature (Lond.) *184*, 903.

ISEKI, S., FUROKAWA, K. and YAMAMOTO, S. (1959). B substance-decomposing enzyme produced by an anaerobic bacterium. II. Chemical action of the B-decomposing enzyme. Proc. Japan Acad. *35*, 513.

ISEKI, S. and MASAKI, S. (1953). Transformation of blood group substance by bacterial enzyme. Proc. Japan Acad. *29*, 460.

ISSELBACHER, K. J. ANDERSON, E. P., KURAHASHI, K. and KALCKAR, H. M. (1956). Congenital galactosaemia: a single enzymatic block in galactose metabolism. Science *123*, 635.

ITANO, H. A. (1957). The human haemoglobins: their properties and genetic control. Advan. Protein Chem. *12*, 216.

ITANO, H. A. (1965). The synthesis and structure of normal and abnormal haemoglobins. *In:* Abnormal haemoglobins in Africa, ed. J. H. P. Jonxis. Blackwell, Oxford.

ITANO, H. A. (1966). Genetic regulation of peptide synthesis in haemoglobins. J. Cell Physiol. *67*, suppl. 1, 65.

ITANO, H. A. and NEEL, J. V. (1950). A new inherited abnormality of human haemoglobin. Proc. Natl. Acad. Sci. U.S. *36*, 613.

ITANO, H. A. and ROBINSON, E. A. (1960). Genetic control of the α-chain and β-chains of hemoglobin. Proc. Natl. Acad. Sci. U.S. *46*, 1492.

JACOB, F. and MONOD, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318.

JATZKEWITZ, H. and MEHL, E. (1969). Cerebroside-sulphatase and arylsulphatase A deficiency in metachromatic leukodystrophy (ML). J. Neurochem. *16*, 19.

JENNE, J. W. (1965). Partial purification and properties of the isoniazid trans-acetylase in the human liver. Its relationship to the acetylation of p-aminosalicylic. J. Clin. Invest. *44*, 1992.

JEPSON, J. B. (1965). Hartnup disease. *In:* The metabolic basis of inherited disease, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. McGraw-Hill, New York.

JERVIS, G. A. (1960). Detection of heterozygotes for phenylketonuria. Clin. Chim. Acta *5*, 471.

JERVIS, G. A. (1950). Excretion of phenylalanine and derivatives in phenylpyruvic oligophrenia. Proc. Soc. Exptl. Biol. Med. *75*, 83.

JERVIS, G. A. (1953). Phenylpyruvic oligophrenia: deficiency of phenylalanine oxidising system. Proc. Soc. Exptl. Biol. Med. *82*, 514.

JOHNSON, F. M., KANAPI, C. G., RICHARDSON, R. H., WHEELER, M. R. and STONE, W. S. (1966). An analysis of polymorphisms among isozyme loci in dark and light Drosophila ananassae strains from American and Western Samoa. Proc. Natl. Acad. Sci. U.S. *56*, 119.

JONES, R. T., BRIMHALL, B., HUISMAN, T. H. J., KLEIHAUER, E. and BETKE, K. (1966). Hemoglobin Freiburg: abnormal hemoglobin due to deletion of a single aminoacid residue. Science *154*, 1024.

JONES, R. T., SCHROEDER, W. A., BALOG, J. E. and VINOGRAD, J. R. (1959). Gross structure of hemoglobin H. J. Am. Chem. Soc. *81*, 3161.

JUSTICE, P., O'FLYNN, M. E. and HSIA, D. Y. Y. (1967). Phenylalanine hydroxylase activity in hyperphenylalanemia. Lancet *1*, 928.

KABAT, E. A. (1956). Blood group substances; their chemistry and immunochemistry. Academic Press, New York.

KABAT, E. A. and LESKOWITZ, S. (1955). Immunochemical studies on blood groups. XVII. Structural units involved in blood group A and specificity. J. Am. Chem. Soc. *77*, 5159.

KALCKAR, H. M., ANDERSON, E. P. and ISSELBACHER, K. J. (1956). Galactosemia, a congenital defect in a nucleotide transferase. Biochim. Biophys. Acta *20*, 262.

KALOW, W. (1959). Cholinesterase types. *In:* Biochemistry of human genetics, Ciba Foundation Symp., ed. G. E. W. Wolstenholme and C. M. O'Connor. Churchill, London.

KALOW, W. and DAVIES, R. O. (1958). The activity of various esterase inhibitors towards atypical human serum cholinesterase. Biochem. Pharmacol. *1*, 183.

KALOW, W. and GENEST, K. (1957). A method for the detection of atypical forms of human serum cholinesterases. Determination of dibucaine numbers. Canad. J. Biochem. Physiol. *35*, 339.

KALOW, W. and STARON, N. (1957). On the distribution and inheritance of atypical forms of human serum cholinesterase as indicated by dibucaine numbers. Canad. J. Biochem. Physiol. *35*, 1305.

KAMARYT, J. and LAXOVA, R. (1965). Amylase heterogeneity. Some genetic and clinical aspects. Humangenetik *1*, 579.

KAPLAN, J.-C. and BEUTLER, E. (1967). Electrophoresis of red cell NADH- and NADPH-diaphorases in normal subjects and patients with congenital methemoglobinemia. Biochem. Biophys. Res. Commun. *29*, 605.

KAPLAN, N. O., EVERSE, J. and ADMIRAAL, J. (1968). Significance of substrate inhibition of dehydrogenases. Ann. N.Y. Acad. Sci. *151*, 400.

KARP, G. W. JR. and SUTTON, H. E. (1967). Some new phenotypes of human red cell acid phosphatase. Am. J. Hum. Genet. *19*, 54.

KAZIRO, K., KIKUCHI, G., NAKAMURA, H. and YOSHIYA, M. (1952). Die Frage nach der physiologischen Funktion der Katalase im menschlichen Organismus: Notiz über die Entdeckung einer Konstitutionsanomalie 'Anenzymia catalasea'. Chem. Ber. *85*, 866.

KEITT, A. S. (1966). Pyruvate kinase deficiency and related disorders of red cell glycolysis. Am. J. Med. *41*, 742.

KELLEY, W. N. (1968). Hypoxanthine-guanine phosphoribosyl transferase deficiency in the Lesch–Nyhan syndrome and gout. Federation Proc. *27*, 1047.

KELLEY, W. N., ROSENBLOOM, F. M., HENDERSON, J. F. and SEEGMILLER, J. E. (1967). A specific enzyme defect in gout associated with overproduction of uric acid. Proc. Natl. Acad. Sci. U.S. *57*, 1735.

KIIL, R. and ROKKONES, T. (1964). Late manifesting variant of branched-chain ketoaciduria (maple syrup urine disease). Acta Pediat. *53*, 356.

KIMURA, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. Camb. *11*, 247.

KING, J. L. and JUKES, T. H. (1969). Non-Darwinian evolution. Science *164*, 788.

KIRK, R. L. (1968). The haptoglobin groups in man. Karger, Basel.

KIRKMAN, H. N. (1959). Characteristics of glucose-6-phosphate dehydrogenase from normal and primaquine sensitive erythrocytes. Nature (Lond.) *184*, 1291.

KIRKMAN, H. N. (1968). Glucose-6-phosphate dehydrogenase variants and drug-induced hemolysis. Ann. N.Y. Acad. Sci. *151*, 753.

KIRKMAN, H. N. and BYNUM, E. (1959). Enzymic evidence of a galactosaemic trait in parents of galactosaemic children. Ann. Hum. Genet. Lond. *23*, 117.

KIRKMAN, H. N. and HENDRICKSON, E. M. (1963). Sex-linked electrophoretic difference in glucose-6-phosphate dehydrogenase. Am. J. Hum. Genet. *15*, 241

KIRKMAN, H. N., MCCURDY, P. R. and NAIMAN, J. L. (1964a). Functionally abnormal G6PD. Cold Spring Harbor. Symp. Quant. Biol. *29*, 391.

KIRKMAN, H. N. and RILEY, H. D. JR. (1961). Congenital non-spherocytic haemolytic anaemia. Studies on a family with a qualitative defect in glucose-6-phosphate dehydrogenase. Am. J. Diseases Children 102, 313.

KIRKMAN, H. N., RILEY, H. D. JR. and CROMWELL, B. B. (1960). Different enzymic expressions of mutants of human glucose-6-phosphate dehydrogenase. Proc. Natl. Acad. Sci. N.Y. *46*, 938.

KIRKMAN, H. N., ROSENTHAL, I. M., SIMON, E. R., CARSON, P. E. and BRINSON, A. G. (1964b). 'Chicago 1' variant of glucose-6-phosphate dehydrogenase in congenital hemolytic disease. J. Lab. Clin. Med. *63*, 715.

KIRKMAN, H. N., SCHETTINI, F. and PICKARD, B. M. (1964c). Mediterranean variant of glucose-6-phosphate dehydrogenase. J. Lab. Clin. Med. *63*, 726.

KIRKMAN, H. N., SIMON, E. R. and PICKARD, B. M. (1965). Seattle variant of glucose-6-phosphate dehydrogenase. J. Lab. Clin. Med. *66*, 834.

KITTO, G. B., WASSARMAN, P. G. and KAPLAN, N. O. (1966). Enzymatically active conformers of mitochondrial malate dehydrogenase. Proc. Natl. Acad. Sci. U.S. *56*, 578.

KLEIHAUER, E. F., REYNOLDS, C. A., DOZY, A. M., WILSON, J. B., MOORES, R. R., BERENSON,

M. P., WRIGHT, C. S. and HUISMAN, T. H. J. (1968). Haemoglobin Bibba or $\alpha_2^{136\ Pro}\beta_2$, an unstable α chain abnormal haemoglobin. Biochim. Biophys. Acta *154*, 220.

KNIGHT, R. A., SELIN, M. J. and HARRIS, H. W. (1959). Genetic factors influencing isoniazid blood levels in humans. Trans. Conf. Chemotherap. Tuberc. (St. Louis) *18*, 52.

KNOX, W. E. and MESSINGER, E. (1958). The detection of the metabolic effect of the recessive gene for phenylketonuria. Am. J. Hum. Genet. *10*, 53.

KOBATA, A., GROLLMAN, E. F. and GINSBURG, V. (1968a). An enzymatic basis for blood type A in humans. Arch. Biochem. Biophys. *124*, 609.

KOBATA, A., GROLLMAN, E. F. and GINSBURG, V. (1968b). An enzymatic basis for blood type B in humans. Biochem. Biophys. Res. Commun. *32*, 272.

KOCH, J., SKOKSTAD, E. L. R., WILLIAMS, H. E. and SMITH, L. H. JR. (1967). Deficiency of 2-oxo-glutarate-glyoxylate carboligase activity in primary hyperoxaluria. Proc. Natl. Acad. Sci. U.S. *57*, 1123.

KOMROWER, G. M., SCHWARZ, V., HOLZEL, A. and GOLBERG, L. (1956). A clinical and biochemical study of galactosaemia. Arch. Disease Childhood *31*, 254.

KONOTEY-AHULU, F. I. D., GALLO, E., LEHMANN, H. and RINGELHANN, B. (1968). Haemoglobin Korle-Bu (β73 aspartic acid $\rightarrow$ asparagine). J. Med. Genet. *5*, 107.

KOSCIELAK, J. (1967). Isolation of ABO antigens from red cells. *In:* Methods in immunology and immunochemistry, Vol. 1, ed. C. A. Williams and M. W. Chase. Academic Press, New York.

KRAUS, A. P., LANGSTON, M. F. and LYNCH, B. L. (1968). Red cell phosphoglycerate kinase deficiency. Biochem. Biophys. Res. Commun. *30*, 173.

KRAUS, A. P. and NEELY, C. L. JR. (1964). Human erythrocyte lactate dehydrogenase: four genetically determined variants. Science *145*, 595.

KROOTH, R. S. (1964). Properties of diploid cell strains developed from patients with an inherited abnormality of uridine biosynthesis. Cold Spring Harbor Symp. Quant. Biol. *29*, 189.

KUMAHARA, Y., FEINGOLD, D. S., FREEDBERG, I. M. and HIATT, H. H. (1961). Studies of pentose metabolism in normal subjects and in patients with pentosuria and pentosuria trait. J. Clin. Endocr. *21*, 887.

KUNKEL, H. G., CEPPELLINI, R., MULLER-EBERHARD, U. and WOLF, J. (1957). Observations on the minor basic haemoglobin component in blood of normal individuals and patients with thalassaemia. J. Clin. Invest. *36*, 1615.

LABIE, D., SCHROEDER, W. A. and HUISMAN, T. H. J. (1966). The aminoacid sequence of the $\delta\beta$ chains of haemoglobin Lepore Augusta = Lepore Washington. Biochim. Biophys. Acta *127*, 428.

LABERGE, C. (1969). Hereditary tyrosinemia in a French Canadian isolate. Am. J. Hum. Genet. *21*, 36.

LA DU, B. N. (1966). Alcaptonuria. *In:* The metabolic basis of inherited diseases, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. 2nd edition. McGraw-Hill, New York.

LA DU, B. N. (1967). The enzymatic deficiency in tyrosinemia. Am. J. Diseases Children *113*, 54.

LA DU, B. N. (1967). Histidinemia. Am. J. Diseases Children *113*, 88.

LA DU, B. N., ZANNONI, V. G., LASTER, L. and SEEGMILLER, J. E. (1958). The nature of the defect in tyrosine metabolism in alkaptonuria. J. Biol. Chem. *230*, 251.

LA DU, B. N., HOWELL, R. R., JACOBY, G. A., SEEGMILLER, J. E. and ZANNONI, V. G. (1962). The enzymatic effect in histidinemia. Biochem. Biophys. Res. Comm. *7*, 398.

LAMBOTTE-LEGRAND, J. and LAMBOTTE-LEGRAND, C. (1958). Notes complémentaires sur la drépanocytose. II. Sicklémie et malaria. Ann. Soc. Belge. Med. Trop. *38*, 45.

LANDING, R. H., SILVERMAN, F. N., CRAIG, J. M., JACOBY, M. D., LAHEY, M. E. and CHADWICK, D. L. (1964). Familial neurovisceral lipidosis. An analysis of eight cases of a syndrome previously reported as 'Hurler variant', 'pseudo-Hurler disease' and Tay–Sachs disease with visceral involvement. Am. J. Diseases Children *108*, 503.

LANDSTEINER, K. (1901). Über Agglutinationserscheinungen normalen menschlichen Blutes. Wien. Klin. Wschr. *14*, 1132.

LARIZZA, P., BRUNETTI, P., GRIGNANI, F. and VENTURE, S. (1958). L'individualita bioenzimatica dell'eritrocite 'fabico' sopra alcune anomalie biochemiche ed enzimatiche della emazie nei pazienti affetti da favismo e nei loro familiari. Haematologia *43*, 205.

LASKER, M. (1941). Essential fructosuria. Human Biol. *13*, 51.

LASKER, M., ENKLEWITZ, M. and LASKER, G. W. (1936). The inheritance of L-xyloketosuria (essential pentosuria). Human Biol. *8*, 243.

LASTER, L., MUDD., S. H., FINKELSTEIN, J. D. and IRREVERRE, F. (1965). Homocystinuria due to cystathionine synthase deficiency: the metabolism of L-methionine. J. Clin. Invest. *44*, 1708.

LAYZER, R. B., ROWLAND, L. P. and RANNEY, H. M. (1967). Muscle phosphofructokinase deficiency. Arch. Neurol. *17*, 512.

LEGUM, C. P. and NITOWSKY, H. M. (1969). Studies on leucocyte brancher enzyme activity in a family with type IV glycogenosis. J. Pediat. *74*, 84.

LEHMANN, H. and CARRELL, R. W. (1969). Variations in the structure of human haemoglobin. Brit. Med. Bull. *25*, 14.

LEHMANN, H. and HUNTSMAN, R. G. (1966). Man's haemoglobins. North-Holland Publishing Co., Amsterdam.

LEHMANN, H., HUNTSMAN, R. G. and AGER, J. A. M. (1966). The haemoglobinopathies and thalassaemia. *In:* The metabolic basis of inherited disease, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. 2nd edition. McGraw-Hill, New York.

LEHMANN, H. and RYAN, E. (1956). The familial incidence of low pseudocholinesterase level. Lancet *2*, 124.

LEHRS, H. (1930). Über die gruppenspezifische Eigenschaften des menschlichen Speichels. Z. Immun. Forsch. *66*, 175.

LEVINE, P., ROBINSON, E., CELANO, M., BRIGGS, O. and FALKINBURG, L. (1955). Gene interaction resulting in suppression of blood group substance B. Blood *10*, 1100.

LEWIS, E. B. (1951). Pseudo allelism and gene evolution. Cold Spring Harbor Symp. Quant. Biol. *16*, 159.

LEWIS, G. M., SPENCER-PEET, J. and STEWART, K. M. (1963). Infantile hypoglycaemia due to inherited deficiency of glycogen synthetase in liver. Arch. Disease Childhood *38*, 40.

LEWIS, G. M., STEWART, K. M. and SPENCER-PEET, J. (1962). Absence of the liver enzyme, UDPG-glycogen 1, 4-transglycosylase, as a cause of infantile hypoglycaemia. Biochem. J. *84*, 115P.

LEWIS, W. H. P., CORNEY, G. and HARRIS, H. (1968). Pep A 5–1 and Pep A 6–1: two new variants of peptidase A with features of special interest. Ann. Hum. Genet. Lond. *32*, 35.

LEWIS, W. H. P. and HARRIS, H. (1967). Human red cell peptidases. Nature (Lond.) *215*, 351.

LEWIS, W. H. P. and HARRIS, H. (1969a). *In vitro* 'hybridisation' of peptidase A subunits between different human phenotypes, and between human and monkey. Ann. Hum. Genet. Lond. *33*, 89.

LEWIS, W. H. P. and HARRIS, H. (1969b). Peptidase D (prolidase) variants in man. Ann. Hum. Genet. Lond. *32*, 317.

LEWONTIN, R. C. (1967). An estimate of average heterozygosity in man. Amer. J. Hum. Genet. *19*, 681.

LEWONTIN, R. C. and HUBBY, J. L. (1966). A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural population of Drosophila pseudoobscura. Genetics *54*, 595.

LEVIN, B. (1967). Arginino succinic aciduria. Am. J. Diseases Children *113*, 162.

LESCH, M. and NYHAN, W. L. (1964). A familial disorder of uric acid metabolism and central nervous system function. Am. J. Med. 36, 561.

LIDDELL, J., BROWN, D., BEALE, D., LEHMANN, H. and HUNTSMAN, R. G. (1964). A new haemoglobin – J_α Oxford found during a survey of an English population. Nature (Lond.) *204*, 269.

LIDDELL, J., LEHMANN, H. and DAVIES, D. (1963). Harris and Whittaker's pseudocholinesterase variant with increased resistance to fluoride. A study of four families and the identification of the homozygote. Acta Genet. *13*, 95.

LIDDELL, J., LEHMANN, H. and SILK, E. (1962). A 'silent' pseudocholinesterase gene. Nature (Lond.) *193*, 561.

LIE-INJO, L. E., LIE, H. G., AGER, J. A. M. and LEHMANN, H. (1962). α-thalassaemia as a cause of hydrops foetalis. Brit. J. Haematol. *8*, 1.

LIFSCHITZ, F. (1966). Congenital lactase deficiency. J. Pediat. *69*, 229.

LINDER, D. and GARTLER, S. M. (1965a). Distribution of glucose-6-phosphate dehydrogenase electrophoretic variants in different tissues of heterozygotes. Am. J. Hum. Genet. *17*, 212.

LINDER, D. and GARTLER, S. M. (1965b). Glucose-6-phosphate dehydrogenase mosaicism: utilisation as a cell marker in the study of leiomyomas. Science *150*, 67.

LIVINGSTONE, F. B. (1967). Abnormal hemoglobins in human populations. Aldine Publishing Co., Chicago.

LLOYD, K. O., KABAT, E. A., LAYUG, E. J. and GRUEZO, F. (1966). Immunochemical studies on blood groups. 34. Structure of some oligosaccharides produced by alkaline degradation of blood group A, B, and H substances. Biochemistry 5, 1489.

LLOYD, K. O., KABAT, E. A. and LICERIO, E. (1968). Immunochemical studies on blood groups. 38. Structure and activities of oligosaccharides produced by alkaline degradation of blood group Lewis^a substance. Proposed structure of the carbohydrate chains of human blood group A, B, H, Le^a and Le^b substances. Biochemistry *7*, 2976.

LÖHR, G. W. and WALLER, H. D. (1962). Eine neue enzymopenische hämolytische Anämie mit Glutathionreductase-Mangel. Med. Klin. *57*, 1521.

LONG, W. K. (1967). Glutathione reductase in red blood cells: variant associated with gout. Science *155*, 712.

LONG, W. K., KIRKMAN, H. N., SUTTON, H. E. (1965). Electrophoretically slow variants of glucose-6-phosphate dehydrogenase from red cells of Negroes. J. Lab. Clin. Med. *65*, 81.

LUFFMAN, J. E. and HARRIS, H. (1967). A comparison of some properties of human red cell acid phosphatase in different phenotypes. Ann. Hum. Genet. Lond. *30*, 387.

LUZZATTO, L. and ALLAN, N. C. (1965). Different properties of glucose-6-phosphate dehydrogenase from human erythrocytes with normal and abnormal enzyme levels. Biochem. Biophys. Res. Commun. *21*, 547.

LUZZATTO, L., USANGA, E. A. and REDDY, S. (1969). Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. Science *164*, 839.

LYON, M. F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. Am. J. Hum. Genet. *14*, 135.

MACIVER, J. E., WENT, L. N. and IRVINE, R. A. (1961). Hereditary persistence of foetal haemoglobin: a family study suggesting allelism of the F gene to the S and C haemoglobin genes. Brit. J. Haematol. *7*, 373.

MACKENZIE, D. Y. and WOOLF, L. I. (1959). Maple syrup urine disease – an inborn error of the metabolism of valine, leucine and isoleucine associated with gross mental deficiency. Brit. med. J. *1*, 90.

MCARDLE, R. B. (1951). Myopathy due to a defect in muscle glycogen breakdown. Clin. Sci. *10*, 13.

MCCARTHY, C. F., BORLAND, J. L. JR., LYNCH, H. J. JR., OWEN, E. E. and TYOR, M. P. (1964). Defective uptake of basic aminoacids and L-cystine by intestinal mucosa of patients with cystinuria. J. Clin. Invest. *43*, 1518.

MCCONNELL, R.B. (1966). The genetics of gastro-intestinal disorders. Oxford University Press.

MCCURDY, P. R., KIRKMAN, H. N., NAIMAN, J. L., JIM, R. T. S. and PICKARD, B. M. (1966). A Chinese variant of glucose-6-phosphate dehydrogenase. J. Lab. Clin. Med. *67*, 374.

MCCURDY, P. R., PEARSON, H. and GERALD, P. S. (1961). A new haemoglobinopathy of unusual genetic significance. J. Lab. Clin. Med. *58*, 86.

MCKUSICK, V. A. (1966). Mendelian inheritance in man. Catalogs of autosomal Dominant, Autosomal recessive, and X linked phenotypes. Johns Hopkins Press, Baltimore.

MCMURRAY, W. C., RATHBUN, J. C., MOHYUDDIN, F. and KOEGLER, S. J. (1963). Citrullinuria. Pediatrics *32*, 347.

MARCUS, D. M. and CASS, L. E. (1969). Glycosphingolipids with Lewis blood group activity: uptake by human erythrocytes. Science *164*, 553.

MARKERT, C. L. (1963). Lactate dehydrogenase isozymes: dissociation and recombination of subunits. Science *140*, 1329.

MARKERT, C. L. (1968). The molecular basis for isozymes. Ann. N.Y. Acad. Sci. *151*, 14.

MARKERT, C. L. and MOLLER, F. (1959). Multiple forms of enzymes: tissue, ontogenetic, and species specific patterns. Proc. Natl. Acad. Sci. U.S. *45*, 753.

MARKS, P. A. (1958). Red cell glucose-6-phosphate and 6-phosphogluconic dehydrogenases and nucleoside phosphorylase. Science *127*, 1338.

MARKS, P. A., BANKS, J. and GROSS, R. T. (1962). Genetic heterogeneity of glucose-6-phosphate dehydrogenase deficiency. Nature (Lond.) *194*, 454.

MARKS, P. A., GROSS, R. T. and HURWITZ, R. E. (1959). Gene action in erythrocyte deficiency of glucose-6-phosphate dehydrogenase deficiency: tissue enzyme levels. Nature (Lond.) *183*, 1266.

MARKS, P. A., SZEINBERG, A. and BANKS, J. (1961). Erythrocyte glucose-6-phosphate de-

hydrogenase of normal and mutant human subjects. Properties of purified enzymes. J. Biol. Chem. *236*, 10.

MARR, A. M. S., DONALD, A. S. R., WATKINS, W. M. and MORGAN, W. T. J. (1967). Molecular and genetic aspects of human blood-group Le^b specificity. Nature (Lond.) *215*, 1345.

MATHAI, C. K. and BEUTLER, E. (1966). Electrophoretic variation of galactose-1-phosphate uridyltransferase. Science *154*, 1179.

MEHL, E. and JATZKEWITZ, M. (1965). Evidence for a genetic block in metachromatic leukodystrophy. Biochem. Biophys. Res. Commun. *19*, 407.

MELARTIN, L. and BLUMBERG, B. S. (1966). Albumin Naskapi: a new variant of serum albumin. Science *153*, 3744.

MIGEON, B. R., DER KALOUSTIAN, V. M., NYHAN. W, L., YOUNG, W. J. and CHILDS, B. (1968). X-linked hypoxanthine-guanine phosphoribosyl transferase deficiency: heterozygote has two clonal populations. Science *160*, 425.

MILLER, A. L. and MCLEAN, P. (1967). Urea cycle enzymes in the liver of a patient with argininosuccinic aciduria. Clin. Sci. *32*, 385.

MILNE, M. D., ASATOOR, A. M., EDWARDS, K. G. D. and LOUGHBRIDGE, L. W. (1961). The intestinal absorption defect in cystinuria. Gut *2*, 323.

MISHU, M. K. and NANCE, W. E. (1969). Further evidence for close linkage of the Hb^{β} and Hb^{δ} loci in man. J. Med. Genet. *6*, 190.

MITOMA, C., AULD, R. M. and UDENFRIEND, S. (1957). On the nature of enzymatic defect in phenylpyruvic oligophrenia. Proc. Soc. Exptl. Biol. N.Y. *94*, 634.

MIYAJI, T., IUCHI, I., SHIBATA, S., TAKEDA, I. and TAMURA, A. (1963). Possible aminoacid substitution in the α chain (α 87 Tyr) of Hb M Iwata. Acta Haematol. Japan *26*, 538.

MIYAMOTO, M. and FITZPATRICK, T. B. (1957). Competitive inhibition of mammalian tyrosinase by phenylalanine and its relationship to hair pigmentation in phenylketonuria. Nature (Lond.) *179*, 199.

MOHYUDDIN, F., RATHBUN, J. C. and MCMURRAY, W. C. (1967). Studies on aminoacid metabolism in citrullinuria. Amer. J. Diseases Children *113*, 152.

MONN, E. (1969a). Relation between blood cell phosphoglucomutase isoenzymes and age of cell population. Scand. J. Haematol. *6*

MONN, E. (1969b). Chromatographic studies on human red cell phosphoglucomutase. Protein Res. *1*, 1.

MORGAN, W. T. J. (1967). Soluble blood group specific substances. *In:* Methods in immunology and immunochemistry p. 75., ed. C. A. Williams and M. W. Chase. Vol. 1, Academic Press, New York.

MORGAN, W. T. J. and VAN HEYNINGEN, R. (1944). The occurrence of A, B and O blood group substances in pseudomucinous ovarian cyst fluids. Brit. J. Exptl. Pathol. *25*, 5.

MORGAN, W. T. J. and WATKINS, W. M. (1948). The detection of a product of the blood group O gene and the relationship of the so-called O substance to the agglutinogens A and B. Brit. J. Exptl. Pathol. *29*, 159.

MORGAN, W. T. J. and WATKINS, W. M. (1953). The inhibition of the haemagglutinins in plant seeds by human blood-group substances and simple sugars. Brit. J. Exptl. Pathol. *34*, 94.

MORGAN, W. T. J. and WATKINS, W. M. (1956). The product of the human blood group A and B genes in individuals belonging to group AB. Nature (Lond.) *177*, 521.

MORGAN, W. T. J. and WATKINS, W. M. (1969). Genetic and biochemical aspects of human blood-group A-, B-, H-, Le^a- and Le^b- specificity. Brit. Med. Bull. *25*, 30.

MORGANTI, G., BEOLCHINI, P. E., VIERUCCI, A. and BÜTLER, R. (1967). Contributions to the genetics of the serum β-lipoproteins in man. I. Frequency, transmission and penetrance of factors Ag(x) and Ag(y). Humangenetik *4*, 262.

MORRIS, M. D., FISHER, D. A. and FISER, R. (1966). Late onset branched chain ketoaciduria (maple syrup urine disease). Journal-Lancet *86*, 149.

MORROW, G. (1967). Citrullinemia. Am. J. Diseases Children *113*, 157.

MORROW, III, G., BARNES, L. A., CARDINALE, G. J., ABELES, R. H., and FLAKS, J. G. (1969). Congenital methylmalonic acidemia: enzymatic evidence for two forms of the disease. Proc. Natl. Acad. Sci. U.S. *63*, 191.

MORTON, N. E. (1964). Genetic studies of Northeastern Brazil. Cold Spring Harbor Symp. Quant. Biol. *29*, 69.

MOTULSKY, A. G. (1964). Current concepts of the genetics of the thalassaemias. Cold Spring Harbor Symp. Quant. Biol. *29*, 399.

MOURANT, A. E. (1946). A 'new' human blood group antigen of frequent occurrence. Nature (Lond.) *158*, 237.

MUDD, S. H., FINKELSTEIN, J. D., IRREVERRE, F. and LASTER, L. (1964). Homocystinuria: an enzymatic defect. Science *143*, 1443.

MUDD, S. H., IRREVERRE, F. and LASTER, L. (1967). Sulfite oxidase deficiency in man: demonstration of the enzymatic defect. Science *156*, 1599.

MULLER, C. J. and KINGMA, S. (1961). Haemoglobin Zurich $\alpha_2^{A}\ \beta_2^{63Arg}$. Biochim. Biophys. Acta *50*, 595.

MURAYAMA, M. (1966). Molecular mechanism of red cell 'sickling'. Science *153*, 145.

MURRAY, P., THOMSON, J. A., MCGIRR, E. M., WALLACE, T. J., MACDONALD, E. M. and MACCABAE, H. J. (1965). Absent and defective iodotyrosine deiodination. Lancet *1*, 183.

NAKAO, K., WADA, O., KITAMURA, T. and UONO, M. (1966). Activity of amino-laevulinic acid synthetase in normal and porphyric human livers. Nature (Lond.) *210*, 838.

NAKAJIMA, H. (1963). Studies on heme α-methenyl oxygenase, a new enzyme which is capable of transforming haemoglobin-haptoglobin to a possible precursor of biliverdin. Proc. 9th Congr. Eur. Soc. Haematology, Lisbon. p. 840. Karger, Basel.

NAKAJIMA, H., TAKAMURA, T., NAKAJIMA, O. and YAMAOKA, K. (1963). Studies on heme α-methenyl oxygenase. J. Biol. Chem. *238*, 3784.

NANCE, W. E. (1967). Genetic studies of human serum and erythrocyte polymorphisms. Ph. D. Thesis. University of Wisconsin.

NANCE, W. E., CLAFLIN, A. and SMITHIES, O. (1963). Lactic dehydrogenase: genetic control in man. Science *142*, 1075.

NANCE, W. E. and SMITHIES, O. (1963). New haptoglobin alleles: a prediction confirmed. Nature (Lond.) *198*, 869.

NANCE, W. E. and UCHIDA, I. (1964). Turner's syndrome, twinning, and an unusual variant of glucose-6-phosphate dehydrogenase. Am. J. Hum. Genet. *16*, 380.

NECHELES, T. F., MALDONADO, N., BARQUET-CHEDIAK, A. and ALLEN, D. M. (1969). Homozygous erythrocyte glutathione-peroxidase deficiency: clinical and biochemical studies. Blood *33*, 164.

NEEB, H., BEIBOER, J. L., JONXIS, J. H. P., KAARS SIJPESTEIJN, J. A. and MULLER, C. J. (1961).

Homozygous Lepore haemoglobin disease appearing as thalassaemia major in two Papuan siblings. Trop. Geogr. Med. *13*, 207.

NEEL, J. V. (1949). The inheritance of sickle cell anaemia. Science *110*, 64.

NEEL, J. V. (1951). The inheritance of the sickling phenomenon, with particular reference to sickle cell disease. Blood *6*, 389.

NEEL, J. V. and SALZANO, F. M. (1967). Further studies on the Xavante Indians. X. Some hypotheses – generalizations resulting from these studies. Am. J. Hum. Genet. *19*, 554.

NG, W. G., DONNELL, G. N. and BERGREN, W. R. (1968). Deficiency of erythrocyte nicotinamide adenine dinucleotide nucleosidase (NADase) activity in the Negro. Nature (Lond.) *217*, 64.

NISHIMURA, E. T., HAMILTON, H. B., KOBARA, T. Y., TAKAHARA, S., OGURA, Y. and DOI, K. (1959). Carrier state in human acatalasemia. Science *130*, 3371.

NORUM, K. R. and GJONE, E. (1967a). Familial serum-cholesterol esterification failure. A new inborn error of metabolism. Biochim. Biophys. Acta *144*, 698.

NORUM, K. R. and GJONE, E. (1967b). Familial plasma lecithin: cholesterol acyltransferase deficiency. Biochemical study of a new inborn error of metabolism. Scand. J. Clin. Lab. Invest. *20*, 231.

NYHAN, W. L., PESEK, J., SWEETMAN, L., CARPENTER, D. G. and CARTER, C. H. (1967). Genetics of an X-linked disorder of uric acid metabolism and cerebral function. Pediat. Res. *1*, 5.

NYMAN, M. (1959). Serum haptoglobin. Methodological and clinical studies. Scand. J. Clin. Lab. Invest. *11*, suppl. *39*.

OATES, J. A., NIRENBERG, P. Z., JEPSON, J. B., SJOERDSMA. A. and UDENFRIEND, S. (1963). Conversion of phenylalanine to phenylethylamine in patients with phenylketonuria. Proc. Soc. Exptl. Biol. Med. *112*, 1078.

OBERHOLZER, V. G., LEVIN, B., BURGESS, E. A. and YOUNG, W. F. (1967). Methylmalonic aciduria. Inborn error of metabolism leading to metabolic acidosis. Arch. Dis. Childhood *42*, 492.

ÖCKERMAN, P. A. and KÖHLIN, P. (1968). Tissue acid hydrolase activities in Gaucher's disease. Scand. J. Clin. Lab. Invest. *22*, 62.

OKADA, S. and O'BRIEN, J. S. (1968). Generalized gangliosidosis: β-galactosidase deficiency. Science *160*, 1002.

OPFELL, R. W., LORKIN, P. A. and LEHMANN, H. (1968). Hereditary non-spherocytic haemolytic anaemia with post-splenectomy inclusion bodies and pigmenturia caused by an unstable haemoglobin Santa Ana-β88 (F4) leucine $\rightarrow$ proline. J. Med. Genet. *5*, 292.

OPITZ, J. M., STILES, F. C., WISE, D., RACE, R. R., SANGER, R., VON GEMMINGEN, G. R., KIERLAND, R. R. CROSS, E. G. and DE GROTT, W. P. (1965). The genetics of angiokeratoma corporis diffusum (Fabry's disease) and its linkage relations with the Xg locus. Amer. J. Hum. Genet. *17*, 325.

PAGLIA, D. E., HOLLAND, P., BAUGHAN, M. A. and VALENTINE, W. N. (1969). Occurrence of defective hexosephosphate isomerization in human erythrocytes and leucocytes. New Engl. J. Med. *280*, 66.

PAGLIA, D. E., VALENTINE, W. N., BAUGHAN, M. A., MILLER, D. R., REED, C. F. and MCINTYRE, O. R. (1968). An inherited molecular lesion of erythrocyte pyruvate kinase. J. Clin. Invest. *47*, 1929.

PAINTER, T. J., WATKINS, W. M. and MORGAN, W. T. J. (1962). Isolation of a B-specific disaccharide from human blood-group B substance. Nature (Lond.) *193*, 1042.

PARE, C. M. B., SANDLER, M. and STACEY, R. C. (1957). 5-hydroxytryptamine deficiency in phenylketonuria. Lancet *1*, 511.

PARKER, W. C. and BEARN, A. G. (1961). Haptoglobin and transferrin variation in humans and primates: two new transferrins in Chinese and Japanese populations. Ann. Hum. Genet. Lond. *25*, 227.

PARR, C. W. (1966). Erythrocyte phosphogluconate dehydrogenase polymorphism. Nature (Lond.) *210*, 487.

PARRINGTON, J. M., CRUICKSHANK, G., HOPKINSON, D. A., ROBSON, E. B. and HARRIS, H. (1968). Linkage relationships between the three phosphoglucomutase loci PGM_1, PGM_2 and PGM_3. Ann. Hum. Genet. Lond. *32*, 27.

PATRICK, A. D. (1965). A deficiency of glucocerebrosidase in Gaucher's disease. Biochem. J. *97*, 17c.

PAULING, L., ITANO, H. A., SINGER, S. J. and WELLS, I. C. (1949). Sickle cell anaemia, a molecular disease. Science *110*, 543.

PATRICK, A. D. and LAKE, B. D. (1969). Deficiency of an acid lipase in Wolman's disease. Nature (Lond.) *222*, 1067.

PENHOET, E. E., KOCHMAN, M., VALENTINE, R. and RUTTER, W. J. (1967). The subunit structure of mammalian fructose diphosphate aldolase. Biochemistry *6*, 2940.

PENHOET, E. E., RAJKUMAR, T. V. and RUTTER, W. J. (1966). Multiple forms of fructose diphosphate aldolase in mammalian tissues. Proc. Natl. Acad. Sci. U.S. *56*, 1275.

PENROSE, L. S. (1954). Quelques principes sur la fréquence des gènes et sa stabilité dans les populations humaines. J. Génét. Hum. *3*, 159.

PENROSE, L. S. (1963). Outline of human genetics. 2nd edition. Heinemann, London.

PERLROTH, M. G., TSCHUDY, D. P., MARVER, H. S., BERARD, C. W., ZEIGEL, R. F., RECHCIGL, M. and COLLINS, A. (1966). Acute intermittent porphyria. New morphologic and biochemical findings. Amer. J. Med. *41*, 149.

PERUTZ, M. F. and LEHMANN, H. (1968). Molecular pathology of human haemoglobin. Nature (Lond.) *219*, 902.

PERUTZ, M. F. and MITCHISON, J. M. (1950). State of haemoglobin in sickle-cell anaemia. Nature (Lond.) *166*, 677.

PERUTZ, M. F., MUIRHEAD, H., COX, J. M. and GOAMAN, L. C. G. (1968). Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. Nature (Lond.) *219*, 131.

PETERS, J. H., GORDON, G. R. and BROWN, P. (1965). The relationship between the capacities of human subjects to acetylate isoniazid, sulfanilamide and sulfamethazine. Life Sci. *4*, 99.

PINTO, P. V. C., NEWTON, W. A. JR. and RICHARDSON, K. E. (1966). Evidence for four types of erythrocyte glucose-6-phosphate dehydrogenase from G-6-PD deficient human subjects. J. Clin. Invest. *45*, 823.

PIOMELLI, S., CORASH, L. M., DAVENPORT, D. D., MIRAGLIA, J. and AMBROSI, E. L. (1968). Glucose-6-phosphate dehydrogenase deficiency (types Gd^{A-} and $Gd^{Mediterranean}$); the result of *in vivo* instability of the mutant enzymes. J. Clin. Invest. *47*, 940.

POOTRAKUL, S., WASI, P. and NA-NAKORN, S. (1967a). Haemoglobin Bart's hydrops foetalis in Thailand. Ann. Hum. Genet. Lond. *30*, 293.

POOTRAKUL, S., WASI, P. and NA-NAKORN, S. (1967b). Studies on haemoglobin Bart's

(Hb-γ_4) in Thailand: the incidence and the mechanism of occurrence in cord blood. Ann. Hum. Genet. Lond. *31*, 149.

PORTER, I. H., BOYER, S. H., WATSON-WILLIAMS, E. J., ADAM, A., SZEINBERG, A. and SINISCALCO, M. (1964). Variation of glucose-6-phosphate dehydrogenase in different populations. Lancet *1*, 895.

PRADER, A. and AURICCHIO, S. (1965). Defects of intestinal disaccharide absorption. Ann. Rev. Med. *16*, 345.

PTASHNE, M. (1967). Isolation of the λ phage repressor. Proc. Natl. Acad. Sci. U.S. *57*, 306.

PUSZTAI, A. and MORGAN, W. T. J. (1963). The aminoacid composition of the human blood-group A, B, H and Le^a specific substances. Biochem. J. *88*, 546.

PUTKONEN, T. (1930). Über die gruppenspezifischen Eigenschaften verschiedener Körperflüssigkeiten. Acta Soc. Med. Fenn. 'Duodecim' A, *14*, No. 12.

QUERIDO, A., STANBURY, J. B., KASSENAAS, A. A. H. and MEIJER, J. W. A. (1956). The metabolism of iodotyrosines. III. Di-iodotyrosine deshalogenating activity of human thyroid tissue. J. Clin. Endocr. 16, 1096.

RACE, R. R. and SANGER, R. (1968). Blood groups in man. 5th edition. Blackwell, Oxford.

RACE, C., ZIDERMAN, D. and WATKINS, W. M. (1968). An α-D-galactosyltransferase associated with the blood group B character. Biochem. J. *107*, 733.

RAMOT, B. and BROK, F. (1964). A new glucose-6-phosphate dehydrogenase mutant (Tel-Hashomer mutant). Ann. Hum. Genet. Lond. *28*, 167.

RAMOT, B., FISHER, S., SZEINBERG, A., ADAM, A., SHEBA, C. and GAFNI, D. (1959). A study of subjects with glucose-6-phosphate dehydrogenase deficiency. II. Investigation of leucocyte enzymes. J. Clin. Invest. *38*, 2234.

RANNEY, H. M., JACOBS, A. S., BRADLEY, T. B. and CORDOVA, F. A. (1963). A 'new' variant of haemoglobin A_2 and its segregation in a family with haemoglobin S. Nature (Lond.) *197*, 164.

RAPER, A. B. (1956). Sickling in relation to morbidity from malaria and other diseases. Brit. Med. J. *1*, 965.

RAPER, A. B., GAMMACK, D. B., HUEHNS, E. R. and SHOOTER, E. M. (1960). Four haemoglobins in one individual. A study of the genetic interaction of Hb G and Hb C. Brit. Med. J. *2*, 1257.

RAPLEY, S., ROBSON, E. B. and HARRIS, H. (1967). Data on the incidence, segregation and linkage relations of the adenylate kinase (AK) polymorphism. Ann. Hum. Genet. Lond. *31*, 237.

RASMUSSEN, K. (1968). Phosphorylethanolamine and hypophosphatasia. Danish Med. Bull. *15* (supplement).

RATHBUN, J. C. (1948). Hypophosphatasia, a new development anomaly. Amer. J. Diseases Children *75*, 822.

RATHBUN, J. C., MACDONALD, J. W., ROBINSON, H. M. and WANKLIN, J. M. (1961). Hypophosphatasia: a genetic study. Arch. Disease Childhood *36*, 540.

REGE, V. P., PAINTER, T. J., WATKINS, W. M. and MORGAN, W. T. J. (1964a). Isolation of serologically active fucose-containing oligosaccharides from human blood-group H substance. Nature (Lond.) *203*, 360.

REGE, V. P., PAINTER, T. J., WATKINS, W. M. and MORGAN, W. T. J. (1964b). Isolation of a serologically active, fucose-containing trisaccharide from human blood-group Le^a substance. Nature (Lond.) *204*, 740.

RENWICK, J. H. (1969). Progress in mapping human autosomes. Brit. Med. Bull. *25*, 65.

REY, J. and FRÉZAL, J. (1967). Les anomalies des disaccharidases. Arch. Franc. Pediat. *24*, 65.

REY, J., FRÉZAL, J., ROYER, P. and LAMY, M. (1966). L'absence congénitale de lipase pancréatique. Arch. Franc. Pediat. *23*, 5.

RIEDER, R. F. and BRADLEY, T. B. (1968). Hemoglobin Gun Hill: an unstable protein associated with chronic hemolysis. Blood *32*, 355.

RIENSKOU, T. (1968). The Gc system. Series Haematologica *1*, 1, 21.

RITTER, H. and WENDT, G. G. (1964). Untersuchung von 223 Familien zur Formalen Genetik des INV-Polymorphismus. Humangenetik *1*, 123.

ROBSON, E. B. and HARRIS, H. (1965). Genetics of the alkaline phosphatase polymorphism of the human placenta. Nature (Lond.) *207*, 1257.

ROBSON, E. B. and HARRIS, H. (1966). Further data on the incidence and genetics of the serum cholinesterase phenotype C_5+. Ann. Hum. Genet. Lond. *29*, 403.

ROBSON, E. B. and HARRIS, H. (1967). Further studies on the genetics of placental alkaline phosphatase Ann. Hum. Genet. Lond. *30*, 219.

ROPARTZ, C., LENOIR, J. and RIVAT, L. (1961). A new inheritable property of human sera: the InV factor. Nature (Lond.) *189*, 586.

ROSENBERG, L. E. (1966). Cystinuria: genetic heterogeneity and allelism. Science *154*, 1341.

ROSENBERG, L. E., ALBRECHT, I. and SEGAL, S. (1967). Lysine transport in human kidney: evidence for two systems. Science *155*, 1426.

ROSENBERG, L. E., DOWNING, S., DURANT, J. L. and SEGAL, S. (1966). Cystinuria: biochemical evidence for three genetically distinct diseases. J. Clin. Invest, *45*, 365.

ROSENBERG, L. E., LILLJEQVIST, A.-CH. and HSIA, Y. E. (1968). Methylmalonic aciduria. New Engl. J. Med. *278*, 1319.

ROSENBERG, L. E., DURANT, J. L. and ELSAS, L. J. (1968). Familial iminoglycinuria. An inborn error of renal tubular transport. New Engl. J. Med. *278*, 1407.

ROSENBLOOM, F. M., KELLEY, W. N., HENDERSON, J. F. and SEEGMILLER, J. E. (1967). Lyon hypothesis and X-linked disease. Lancet *2*, 305.

RUSSELL, J. D. and DEMARS, R. (1967). UDP-glucose: α-D-galactose-1-phosphate uridylyltransferase activity in cultured human fibroblasts. Biochem. Genet. *1*, 11.

RUSSELL, A., LEVIN, B., OBERHOLZER, V. G. and SINCLAIR, L. (1962). Hyperammonemia. A new instance of an inborn enzymatic defect of the biosynthesis of urea. Lancet *2*, 699.

RUTTER, W. J., RAJKUMAR, T., PENHOET, E. and KOCHMAN, M. (1968). Adolase variants: structure and physiological significance. Ann. N.Y. Acad. Sci. *151*, 102.

SACHS, B., STERNFELD, L. and KRAUS, G. (1942). Essential fructosuria: its pathophysiology. Am. J. Diseases Children *63*, 252.

SALZMANN, J., DEMARS, R. and BENKE, P. (1968). Single-allele expression at an X-linked hyperuricemia locus in heterozygous human cells. Proc. Natl. Acad. Sci. U.S. *60*, 545.

SANTACHIARA, A. S. B. and MODIANO, G. (1969). Ultracentrifuge studies of red cell phosphoglucomutase. Nature (Lond.) *233* 625.

SCHMID, K., TOKITA, K. and YOSHIZAKI, H. (1965). The α_1-acid glycoprotein variants of normal Caucasian and Japanese individuals. J. Clin. Invest. *44*, 1394.

SANSONE, G., CARRELL, R. W. and LEHMANN, H. (1967). Haemoglobin Genova: β28 (B10) leucine → proline. Nature (Lond.) *214*, 877.

DI SANT'AGNESE, P. A., ANDERSEN, D. H. and MASON, H. (1950). Glycogen storage disease of the heart. II. Critical review of the literature. Pediatrics *6*, 607.

SCHAPIRA, F., SCHAPIRA, G. and DREYFUS, J.-C. (1961). La lésion enzymatique de la fructosurie benigne. Enzymol. Biol. clin. *1*, 170.

SCHIFF, F. (1927). Über den serologische Nachweiss der Blutgruppeneigenschaft O. Klin. Wschr. *6*, 303.

SCHIFF, F. and SASAKI, H. (1932). Der Ausscheidungstypus, ein auf serologischen Wege nachweisbares Mendelndes Merkmal. Klin. Wschr. *11*, 1426.

SCHIFFMAN, G., KABAT, E. A. and LESKOWITZ, S. (1962). Immunochemical studies on blood groups. XXVI. The isolation of oligosaccharides from human ovarian cyst blood group A substance including two disaccharides and a trisaccharide involved in the specificity of the blood group A antigenic determinant. J. Am. Chem. Soc. *84*, 73.

SCHIFFMAN, G., KABAT, E. A. and THOMPSON, W. (1964). Immunochemical studies on blood groups. XXX. Cleavage of A, B and H blood-group substances by alkali. Biochemistry *3*, 113.

SCHIFFMAN, G., KABAT, E. A. and THOMPSON, W. (1964). Immunochemical studies on blood groups. XXXII. Immunochemical properties of and possible partial structures for the blood group A, B and H antigenic determinants. Biochemistry *3*, 587.

SCHMID, K., BINETT, J. P., TOKITA, K., MOROZ, L. and YOSHIZAKI, H. (1964). The polymorphic forms of α_1-acid glycoprotein of normal Caucasian individuals. J. Clin. Invest. *43*, 2347.

SCHMID, R., ROBBINS, P. W. and TAUT, R. R. (1959). Glycogen synthesis in muscle lacking phosphorylase. Proc. Natl. Acad. Sci. U.S. *45*, 1236.

SCHIMKE, R. N., MCKUSICK, V. A., HUANG, T. and POLLACK, A. D. (1965). Homocystinuria: studies of 20 families with 38 affected members. J. Am. Med. Ass. *193*, 711.

SCHNEIDER, A. S., DUNN, I., IBSEN, K. H. and WEINSTEIN, I. M. (1968a). Inherited triosephosphate isomerase deficiency. Erythrocyte carbohydrate metabolism and preliminary studies of the erythrocyte enzyme. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

SCHNEIDER, A. S., VALENTINE, W. N., HATTORI, M. and HEINS, H. L. (1965). Hereditary hemolytic anaemia with triosephosphate isomerase deficiency. New Engl. J. Med. *272*, 229.

SCHNEIDER, A. S., VALENTINE, W. N., BAUGHAN, M. A., PAGLIA, D. E., SHORE, N. A. and HEINS, H. L. JR. (1968b). Triose phosphate isomerase deficiency. A multi system inherited disorder. Clinical and genetical aspects. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

SCHREIER, K. and FLAIG, H. (1956). Urinary excretion of indolepyruvic acid in normal conditions and Folling's disease. Klin. Wschr. *34*, 1213.

SCHROEDER, W. A., HUISMAN, T. H. J., SHELTON, J. R., SHELTON, J. B., KLEIHAUER, E. F., DOZY, A. M. and ROBBERSON, B. (1968). Evidence for multiple structural genes for the γ chain of human fetal hemoglobin. Proc. Natl. Acad. Sci. U.S. *60*, 537.

SCHRÖTER, W. (1965). Kongenitale nichtspäracytäre hämolytische Anämie bei 2, 3-Diphosphoglyceratmutase Mangel der Erythrocyten im frühen Säuglingsalter. Klin. Wschr. *43*, 1147.

SCHWARZ, V., GOLBERG, L., KOMROWER, G. M. and HOLZEL, A. (1956). Some disturbances of erythrocyte metabolism in galactosaemia. Biochem. J. *62*, 34.

SCOTT, E. M. (1960). The relation of diaphorase of human erythrocytes to inheritance of methaemoglobinemia. J. Clin. Invest. *39*, 1176.

SCOTT, E. M. (1966). Kinetic comparison of genetically different acid phosphatases of human erythrocytes. J. Biol. Chem. *241*, 3049.

SCOTT, E. M. and GRIFFITH, I. (1959). The enzymic defect of hereditary methaemoglobinaemia: diaphorase. Biochim. Biophys. Acta *34*, 584.

SCRIVER, C. R. (1967). Aminoacid transport in mammalian kidney. *In:* Aminoacid metabolism and genetic variation, Ed. W. L. Nyhan. McGraw Hill, New York.

SCRIVER, C. R. (1968). Renal tubular transport of proline hydroxyproline and glycine. III. Genetic basis for more than one mode of transport in human kidney. J. Clin. Invest. *47*, 823.

SCRIVER, C. R. and HECHTMEN, P. (1970). Human genetics of membrane transport with emphasis on aminoacids. *In:* Advances in human genetics, ed. H. Harris and K. Hirschhorn. vol. 1. Plenum Press, New York.

SCRIVER, C. R., LAROCHELLE, J. and SILVERBERG, M. (1967). Hereditary tyrosinemia and tyrosyluria in a French Canadian geographic isolate. Am. J. Diseases Children *113*, 41.

SEEGMILLER, J. E., ROSENBLOOM, F. M. and KELLEY, W. N. (1967). Enzyme defect associated with a sex-linked human neurological disorder and excessive purine synthesis. Science *155*, 1682.

SHAFER, I. A., SCRIVER, C. R. and EFRON, M. L. (1962). Familial hyperprolinaemia, cerebral dysfunction, and renal anomalies occurring in a family with hereditary nephropathy and deafness. New Engl. J. Med. *267*, 51.

SHELDON, W. (1964). Congenital pancreatic lipase deficiency. Arch. Disease Childhood. *39*, 268.

SHAW, C. R. (1965). Electrophoretic variation in enzymes. Science *149*, 936.

SHAW, C. and BARTO, E. (1963). Genetic evidence for the subunit structure of lactate dehydrogenase isozymes. Proc. Natl. Acad. Sci. U.S. *50*, 211.

SHAW, C. R., SYNER, F. N. and TASHIAN, R. E. (1962). New genetically determined molecular form of erythrocyte esterase in man. Science *138*, 31.

SHEN, L., GROLLMAN, E. F. and GINSBURG, V. (1968). An enzymatic basis for secretor status and blood group substance specificity in humans. Proc. Natl. Acad. Sci. U.S. *59*, 224.

SHIBATA, S., MIYAJI, T., KARITA, K., IUCHI, I., OHBA, Y. and YAMAMOTO, K. (1967). A new type of hereditary nigremia discovered in Akita – hemoglobin $M_{\text{Hyde Park}}$ disease. Proc. Japan Acad. *43*, No. 1.

SHIM, B. S. and BEARN, A. G. (1964). The distribution of haptoglobin subtypes in various populations, including subtype patterns in some non-human primates. Am. J. Hum. Genet. *16*, 477.

SHOWS, T. B. JR., TASHIAN, R. E. and BREWER, G. J. (1964). Erythrocyte glucose-6-phosphate dehydrogenase in Caucasians: new inherited variant. Science *145*, 1056.

SHREFFLER, D. C., BREWER, G. J., GALL, J. C. and HONEYMAN, M. S. (1967). Electrophoretic variation in human serum ceruloplasmin: a new genetic polymorphism. Biochem. Genet. *1*, 101.

SICK, K., BEALE, D., IRVINE, D., LEHMANN, H., GOODALL, P. T. and MACDOUGALL, S. (1967). Haemoglobin $G_{\text{Copenhagen}}$ and Haemoglobin $J_{\text{Cambridge}}$. Two new β-chain variants of haemoglobin A. Biochim. Biophys. Acta *140*, 231.

SIMPSON, E. (1966). Factors influencing cholinesterase activity in a Brazilian population. Am. J. Hum. Genet. *18*, 243.

SIMPSON, N. E. and KALOW, W. (1964). The 'silent' gene for serum cholinesterase. Am. J. Hum. Genet. *16*, 180.

SLOAN, H. R., UHLENDORF, B. W., KANFER, J. N., BRADY, R. O. and FREDRICKSON, D. S. (1969) Deficiencyof sphingomyelin-cleaving enzyme activity in tissue cultures derived from patients with Niemann–Pick disease. Biochem. Biophys. Res. Commun. *34*, 582.

SMITH, E. W. and TORBERT, J. V. (1958). Two abnormal haemoglobins with evidence for a new genetic locus for haemoglobin formation. Bull. Johns Hopkins Hosp. *102*, 38.

SMITH, L. H. JR., SULLIVAN, M. and HUGULEY, C. M. JR. (1961). Pyrimidine metabolism in man. IV. The enzymatic defect of oroticaciduria. J. Clin. Invest. *40*, 656.

SMITHIES, O. (1955). Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. Biochem. J. *61*, 629.

SMITHIES, O. (1957). Variations in human serum β-globulins. Nature (Lond.) *180*, 1482.

SMITHIES, O. (1964). Chromosomal rearrangements and protein structure. Cold Spring Harbor Symp. Quant. Biol. *29*, 309.

SMITHIES, O. and CONNELL, G. E. (1959). Biochemical aspects of the inherited variations in human serum haptoglobins and transferrins. *In:* Ciba Foundation Symposium on Biochemistry of Human Genetics, ed. G. E. Wolstenholme and C. M. O'Connor, Churchill, London.

SMITHIES, O., CONNELL, G. E. and DIXON, G. H. (1962a). Inheritance of haptoglobin subtypes. Am. J. Hum. Genet. *14*, 14.

SMITHIES, O., CONNELL, G. E. and DIXON, G. H. (1962b). Chromosomal rearrangements and the evolution of haptoglobin genes. Nature (Lond.) *196*, 232.

SMITHIES, O., CONNELL, G. E. and DIXON, G. H. (1966). Gene action in the human haptoglobins. I. Dissociation into constituent polypeptide chains. J. Mol. Biol. *21*, 213.

SMITHIES, O. and HILLER, O. (1959). The genetic control of transferrin in humans. Biochem. J. *72*, 121.

SMITHIES, O. and WALKER, N. F. (1955). Genetic control of some serum proteins in normal humans. Nature (Lond.) *176*, 1265.

SMITHIES, O. and WALKER, N. F. (1956). Notation for serum protein groups and the genes controlling their inheritance. Nature (Lond.) *178*, 694.

SNEATH, J. S. and SNEATH, P. H. A. (1955). Transformation of Lewis groups of human red cells. Nature (Lond.) *176*, 172.

SNYDER, L. H. (1959). Fifty years of medical genetics. Science *129*, 7.

SNYDERMAN, S. E. (1967). Maple syrup urine disease. *In:* Amino acid metabolism and genetic variation, ed. W. L. Nyhan. McGraw-Hill, New York.

SNYDERMAN, S. E., NORTON, P. and HOLT, L. E. JR. (1955). Effect of tyrosine administration in phenylketonuria. Federation Proc. *14*, 450.

SOBEL, E. H., CLARK, L. C., FOX, R. P. and ROBINOW, M. (1953). Rickets, deficiency of alkaline phosphatase activity and premature loss of teeth in childhood. Pediatrics *11*, 309.

SPENCER, N., HOPKINSON, D. A. and HARRIS, H. (1964a). Phosphoglucomutase polymorphism in man. Nature (Lond.) *204*, 742.

SPENCER, N., HOPKINSON, D. A. and HARRIS, H. (1964b). Quantitative differences and gene dosage in the human red cell acid phosphatase polymorphism. Nature (Lond.) *201*, 299.

SPENCER, N., HOPKINSON, D. A. and HARRIS, H. (1968). Adenosine deaminase polymorphism in man. Ann. Hum. Genet. Lond. *32*, 9.

STAMATOYANNOPOULOS, G., FRASER, G. R., MOTULSKY, A. G., FESSAS, P., AKRIVAKIS, A. and PAPAYANNOPOULOU, T. (1966). On the familial predisposition to favism. Am. J. Hum. Genet. *18*, 253.

STAMATOYANNOPOULOS, G., YOSHIDA, A., BACOPOULOS, C. and MOTULSKY, A. G. (1967). Athens variant of glucose-6-phosphate dehydrogenase. Science *157*, 831.

STANBURY, J. B. (1966). Familial goitre. *In:* The metabolic basis of inherited disease, ed. J. B. Stanbury, J. B. Wyngaarden and D. S. Fredrickson. 2nd edition. McGraw-Hill, New York.

STEIN, W. H. (1951). Excretion of aminoacids in cystinuria. Proc. Soc. Exptl. Biol. *78*, 705.

STEINBERG, A. G. (1967). Genetic variations in human immunoglobulins. The Gm and Inv types. *In:* Advances in immunogenetics, ed. T. J. Greenwalt. Lippincott, Philadelphia.

STETSON, C. A. JR. (1966). The state of haemoglobin in sickled erythrocyte. J. Exptl. Med. *123*, 341.

STONE, W. S., WHEELER, M. R., JOHNSON, F. M. and KOJIMA, K.-I. (1968). Genetic variation in natural island populations of members of the Drosophila nasuta and Drosophila ananassae subgroups. Proc. Natl. Acad. Sci. U.S. *59*, 102.

SUNAHARA, S., URANO, M. and OGAWA, M. (1961). Genetical and geographical studies on isoniazid inactivation. Science *134*, 1530.

SWEELEY, C. C. and KLIONSKY, B. (1963). Fabry's disease: classification as a sphingolipidosis and partial characterization of a novel glycolipid. J. Biol. Chem. *238*, PC 3148.

SZEINBERG, A., SHEBA, C. and ADAM, A. (1958). Enzymatic abnormality in erythrocytes of a population sensitive to *Vicia faba* or drug induced haemolytic anaemia. Nature (Lond.) *181*, 1256.

SZULMAN, A. E. (1964). The histological distribution of the blood group substances in man as disclosed by immunofluorescence. Part III. The A, B and H antigens in embryos and fetuses from 18mm in length. J. Exptl. Med. *119*, 503.

SZULMAN, A. E. (1966). Chemistry, distribution, and function of blood group substances. Ann. Rev. Med. *17*, 307.

TADA, K., WADA, Y. and ARAKAWA, T. (1967). Hypervalinemia. Arch. Disease Childhood *113*, 64.

TAKAHARA, S. (1952). Progressive oral gangrene, probably due to a lack of catalase in the blood (acatalasaemia). Lancet *2*, 1101.

TAKAHARA, S. (1968). Acatalasemia in Japan. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

TANAKA, K. R. and BEUTLER, E. (1969). Hereditary hemolytic anemia due to glucose-6-phosphate dehydrogenase Torrance: a new variant. J. Lab. Clin. Med. *73*, 657.

TANAKA, K. R., BUDD, M. A., EFRON, M. L. and ISSELBACHER, K. J. (1966). Isovaleric acidemia: a new genetic defect of leucine metabolism. Proc. Natl. Acad. Sci. U.S. *56*, 236.

TANAKA, K. R. and VALENTINE, W. N. (1968). Pyruvate kinase deficiency. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

TANAKA, K. R., VALENTINE, W. N. and MIWA, S. (1962). Pyruvate kinase (PK) deficiency: hereditary non-spherocytic hemolytic anemia. Blood *19*, 267.

TANIGUCHI, K. and GJESSING, L. R. (1965). Studies on tyrosinosis: 2. Activity of the transaminase, para-hydroxyl-phenyl pyruvate oxidase and homogentisic acid oxidase. Brit. Med. J. *1*, 968.

TARUI, S., KONO, N., NASU, T. and NISHIKAWA, M. (1969). Enzymatic basis for the coexistence of myopathy and hemolytic disease in inherited muscle phosphofructokinase deficiency. Biochem. Biophys. Res. Commun. *34*, 77.

TARUI, S., OKUNO, G., IKURA, Y., TANAKA, T., SUDA, M. and NISHIKAWA, M. (1965). Phosphofructokinase deficiency in skeletal muscle. A new type of glycogenosis. Biochem. Biophys. Res. Commun. *19*, 517.

TASHIAN, R. E. (1959). Phenylpyruvic acid as a possible precursor of o-hydroxyphenylacetic acid in man. Science *129*, 1553.

TASHIAN, R. E., PLATO, C. C. and SHOWS, T. B. (1963). Inherited variant of erythrocyte carbonic anhydrase in Micronesians from Guam and Saipan. Science *140*, 53.

TASHIAN, R. E. and SHAW, M. W. (1962). Inheritance of an erythrocyte acetylesterase variant in man. Amer. J. Hum. Genet. *14*, 295.

TASHIAN, R. E., SHREFFLER, D. C. and SHOWS, T. B. (1968). Genetic and phylogenetic variation in the different molecular forms of mammalian erythrocyte carbonic anhydrases. Ann. N.Y. Acad. Sci. *151*, 64.

TEDESCO, T. A. and MELLMAN, W. J. (1967). Argininosuccinate synthetase activity and citrulline metabolism in cells cultured from a citrullinemic subject. Proc. Natl. Acad. Sci. U.S. *57*, 829.

THIER, S. O., FOX, M., SEGAL, S. and ROSENBERG, L. E. (1964). Cystinuria: *in vitro* demonstration of an intestinal transport defect. Science *143*, 482.

THIER, S. O., SEGAL, S., FOX, M., BLAIR, A. and ROSENBERG, L. E. (1965). Cystinuria: defective intestinal transport of dibasic aminoacids and cystine. J. Clin. Invest. *44*, 442.

TOMLINSON, S. and WESTALL, R. G. (1964). Argininosuccinic aciduria. Argininosuccinase and arginase in human blood cells. Clin. Sci. *26*, 261.

TÖNZ, O. (1968). The congenital methemoglobinemias. Bibliotheca Haematologia No. 28. Karger, Basel.

TOUSTER, O. (1959). Pentose metabolism and pentosuria. Am. J. Med. *26*, 724.

TOWNES, P. L. (1965). Trypsinogen deficiency disease. J. Pediat. *66*, 275.

TOWNES, P. L., BRYSON, M. F. and MILLER, G. (1967). Further observations on trypsinogen deficiency disease: report of a second case. J. Pediat. *71*, 220.

TOWNES, P. L. and MORRISON, M. (1962). Investigation of the defect in a variant of hereditary methemoglobinemia. Blood *19*, 60.

TOWNSEND, E. H., MASON, H. H. and STRONG, P. S. (1951). Galactosemia and its relation to Laennec's cirrhosis: review of literature and presentation of six additional cases. Pediatrics *7*, 760.

TSCHUDY, D. P., PERLROTH, M. G., MARVER, H. S., COLLINS, A., HUNTER, G. and RECHCIGL, M. (1965). Acute intermittent porphyria: the first 'overproduction disease' localized to a specific enzyme. Proc. Natl. Acad. Sci. U.S. *53*, 841.

UHLENDORF, B. W. and MUDD, S. H. (1968). Cystathionine synthase in tissue culture from human skin: enzyme defect in homocystinuria. Science *160*, 1007.

VALENTINE, W. N. (1968). Hereditary hemolytic anemias associated with specific erythrocyte enzymopathies. Calif. Med. *108*, 280.

VALENTINE, W. N., OSKI, F. A., PAGLIA, D. E., BAUGHAN, M. A., SCHNEIDER, A. S. and NAIMAN, J. L. (1967). Hereditary hemolytic anemia with hexokinase deficiency. New Engl. J. Med. *276*, 1.

VALENTINE, W. N., SCHNEIDER, A. S., BAUGHAN, M. A., PAGLIA, D. E. and HEINS, H. L. JR. (1966). Hereditary hemolytic anemia with triose-phosphate isomerase deficiency. Am. J. Med. *41*, 27.

VALENTINE, W. N., TANAKA, K. R. and MIWA, S. (1961). A specific erythrocyte glycolytic enzyme defect (pyruvate kinase) in three subjects with congenital non-spherocytic hemolytic anemia. Trans. Ass. Amer. Physicians *74*, 100.

VANDEPITTE, J. M., ZUELZER, W. W., NEEL, J. V. and COLAERT, J. (1955). Evidence concerning the inadequacy of mutation as an explanation of the frequency of the sickle cell gene in the Belgian Congo. Blood *10*, 341.

VANDEPITTE, J. M. (1959). The incidence of haemoglobinoses in the Belgian Congo. *In:* Abnormal haemoglobins, ed. J. H. P. Jonxis and J. F. Delafresnaye. Blackwell, Oxford.

VAN HOOF, F. (1967). Amylo-1, 6- glucosidase activity and glycogen content of the erythrocytes of normal subjects. Patients with glycogen storage disease and heterozygotes. Eur. J. Biochem. *2*, 271.

VAN HOOF, F. and HERS, H. G. (1968). The abnormalities of lysosomal enzymes in mucopolysaccharidoses. Eur. J. Biochem. *7*, 34.

VESELL, E. S. (1965a). Formation of human lactate dehydrogenase isozyme patterns *in vitro*. Proc. Natl. Acad. Sci. U.S. *54*, 111.

VESELL, E. S. (1965b). Genetic control of isozyme patterns in human tissues. *In:* Progress in medical genetics, ed. A. G. Steinberg and A. G. Bearn. Grune and Stratton, New York.

VESELL, E. S. (1968). Introduction. Multiple molecular forms of enzymes. Ann. N.Y. Acad. Sci. *151*, 5.

VON GIERKE, E. (1929). Hepato-nephromegalia glykogenia. Beitr. Pathol. Anat. *82*, 497.

WADA, Y., TAKA, K., MINEGAWA, A., YOSHIDA, T., MORIKAWA, T. and OKAMURA, T. (1963). Idiopathic hypervalinemia. Probably a new entity of inborn error of valine metabolism. Tohoku J. Exptl. Med. *81*, 46.

WALDENSTRÖM, J. (1937). Studien über Porphyrie. Acta Med. Scand. Suppl. 82.

WALDENSTRÖM, J. (1957). The porphyrias as inborn errors of metabolism. Amer. J. Med. *22*, 758.

WALDENSTRÖM, J. and HAEGER-ARONSEN, B. (1967). The porphyrias: a genetic problem. *In:* Progress in medical genetics, ed. A. G. Steinberg and A. G. Bearn. Vol. 5. Grune and Stratton, New York.

WALLACE, H. W., MOLDAVE, K. and MEISTER, A. (1957). Studies on conversion of phenylalanine to tyrosine in phenylpyruvic oligophrenia. Proc. Soc. Exptl. Biol. *94*, 632.

WALLER, H. D. (1968). Glutathione reductase deficiency. *In:* Hereditary disorders of erythrocyte metabolism, ed. E. Beutler. Grune and Stratton, New York.

WALLER, H. D., LOHR, G., ZYSNO, E., GEROK, W., VOSS, D. and STRAUSS, G. (1965). Glutathionreductasemangel mit hämatologischen und neurologischen Störungen. Klin. Wschr. *43*, 413.

WATKINS, W. M. (1956). The appearance of H specificity following the enzymic inactivation of blood-group B substance. Biochem. Soc. *64*, 21P.

WATKINS, W. M. (1966). Blood group substances. Science *152*, 172.

WATKINS, W. M., KOSCIELAK, J. and MORGAN, W. T. J. (1964). The relationship between the specificity of the blood-group A and B substances isolated from erythrocytes and from secretions. Proc. 9th Congr. Int. Soc. Blood Transfusion, Mexico City, 1962. p. 213.

WATKINS, W. M. and MORGAN, W. T. J. (1952). Neutralization of the anti-H agglutinin in eel serum by simple sugars. Nature (Lond.) *169*, 825.

WATKINS, W. M. and MORGAN, W. T. J. (1955a). Inhibition by simple sugars of enzymes which decompose the blood-group substances. Nature (Lond.) *175*, *676*.

WATKINS, W. M. and MORGAN, W. T. J. (1955b). Some observations on the O and H characters of human blood and secretions. Vox Sang. (O.S.) *5*, 1.

WATKINS, W. M. and MORGAN, W. T. J. (1957a). Specific inhibition studies relating to the Lewis blood-group system. Nature (Lond.) *180*, 1038.

WATKINS, W. M. and MORGAN, W. T. J. (1957b). The A and H character of the blood-group substances secreted by persons belonging to group A_2. Acta Genet. Statist. Med. *6*, 521.

WATKINS, W. M. and MORGAN, W. T. J. (1962). Further observations on the inhibition of blood-group specific serological reactions by simple sugars of known structure. Vox Sang. *7*, 129.

WATKINS, W. M., ZARNITZ, M. L. and KABAT, E. A. (1962). Development of H activity by human blood-group B substance treated with coffee bean α-galactosidase. Nature (Lond.) *195*, 1204.

WATSON, J. D. and CRICK, F. H. C. (1953). Genetical implications of the structure of deoxyribosenucleic acid. Nature (Lond.) *171*, 964.

WATTS, R. L. and WATTS, D. C. (1968). The implications for molecular evolution of possible mechanisms of primary gene duplication. J. Theoret. Biol. *20*, 227.

WATTS, R. W. E., ENGLEMAN, K., KLINENBERG, J. R., SEEGMILLER, J. E. and SJOERDSMA (1963). Enzyme defect in a case of xanthinuria. Nature (Lond.) *201*, 395.

WEATHERALL, D. J. (1969). Genetics of the thalassaemias. Brit. Med. Bull. *25*, 24.

WEATHERALL, D. J. (1965). The thalassaemia syndromes. Blackwell Scientific Press, Oxford.

WEATHERALL, D. J., CLEGG, J. B. and NAUGHTON, M. A. (1965). Globin synthesis in thalassaemia: an *in vitro* study. Nature (Lond.) *208*, 1061.

WEATHERALL, D. J., SIGLER, A. T. and BAGLIONI, C. (1962). Four hemoglobins in each of three brothers. Genetic and biochemical significance. Bull. Johns Hopkins Hosp. *III*, 143.

WEIJERS, H. A., VAN DE KAMER, J. H., MOSSELL, D. A. A. and DICKE, W. K. (1960). Diarrhoea caused by deficiency of sugar splitting enzymes. Lancet *2*, 296.

WEITKAMP, L. R., SHREFFLER, D. C., ROBBINS, J. L., DRACHMANN, O., ADNER, P. L., WIEME, R. J., SIMON, N. M., COOKE, K. B., SANDOR, G., WUHRMANN, F., BRAEND, M. and TARNOKY, A. L. (1967). An electrophoretic comparison of serum albumin variants from nineteen unrelated families. Acta Genet. Basel *17*, 399.

WELLS, I. C. and ITANO, H. A. (1951). The ratio of sickle cell anemia haemoglobin to normal haemoglobin in the sicklemics. J. Biol. Chem. *188*, 65.

WELLS, W. W., PITTMAN, T. A. and EGAN, T. J. (1964). The isolation and identification of galactitol from the urine of patients with galactosaemia. J. Biol. Chem. *239*, 3192.

WELLS, W. W., PITTMAN, T. A., WELLS, H. J. and EGAN, T. J. (1965). The isolation and identification of galactitol from the brains of galactosemic patients. J. Biol Chem. *240*, 1002.

WEST, C. A., GOMPERTZ, B. D., HUEHNS, E. R., KESSEL, I. and ASHBY, J. R. (1967). Demonstration of an enzyme variant in a case of congenital methaemoglobinaemia. Brit. Med. J. *4*, 212.

WESTALL, R. G. (1960). Argininosuccinic aciduria: identification and reactions of the abnormal metabolites in a newly described form of mental disease. With some preliminary metabolic studies. Biochem. J. *77*, 135.

WHEELER, J. T. and KREVANS, J. R. (1961). Homozygous state of persistent fetal haemoglobin and interaction of persistent haemoglobin with thalassaemia. Bull. Johns Hopkins Hosp. *109*, 217.

WILLIAMS, H. E., KENDIG, E. M. and FIELD, J. B. (1963). Leukocyte debranching enzyme in glycogen storage disease. J. Clin. Invest. *42*, 656.

WILLIAMS, H. E. and SMITH, L. H. JR. (1968). L-glyceric aciduria. A new genetic variant of primary hyperoxaluria. New Engl. J. Med. *278*, 233.

WILSON, A. C., CAHN, R. D. and KAPLAN, N. O. (1963). Functions of the two forms of lactic dehydrogenase in the breast muscle of birds. Nature (Lond.) *197*, 331.

WHITTAKER, M. (1964). The pseudocholinesterase variants: esterase levels and increased resistance to fluoride. Acta Genet. *14*, 281.

WHITTAKER, M. (1967). The pseudocholinesterase variants. A study of fourteen families selected via the fluoride resistant phenotype. Acta Genet. *17*, 1.

WINSLOW, R. B. and INGRAM, V. M. (1966). Peptide chain synthesis of human hemoglobins A and A_2. J. Biol. Chem. *241*, 1144.

WISE, D., WALLACE, H. J. and JELLINEK, E. H. (1962). Angiokeratoma corporis diffusum: a clinical study of eight affected families. Quart. J. Med. *31*, 177.

WOLMAN, M., STERK, V. V., GATT, S. and FRENKEL, M. (1961). Primary familial xanthematosis with involvement and calcification of the adrenals: report of 2 more cases of siblings of a previously described infant. Pediatrics *28*, 742.

WOOLF, L. I. (1951). Excretion of conjugated phenylacetic acid in phenylketonuria. Biochem. J. *49*, ix.

WOESE, C. R. (1967). The genetic code, Harper and Row, New York.

WORLD HEALTH, ORGANISATION (1967). Technical Report Series No. 366. Standardisation of procedures for the study of glucose-6-phosphate dehydrogenase. Geneva.

WRIGHT, S. (1966). Polyallelic random drift in relation to evolution. Proc. Natl. Acad. Sci. U.S. *55*, 1074.

WRIGHTSTONE, R. N. and HUISMAN, T. H. J. (1968). Qualitative and quantitative studies of sickle cell hemoglobin in homozygotes and heterozygotes. Clin. Chim. Acta *22*, 593.

WUU, K.-D. and KROOTH, R. S. (1968). Dihydroorotic acid dehydrogenase activity of human diploid cell strains. Science *160*, 539.

YČAS, M. (1969). The biological code. North-Holland Publishing Co., Amsterdam.

YOSHIDA, A. (1967). A single amino acid substitution (asparagine to aspartic acid) between normal (B+) and the common Negro variant (A+) of human glucose-6-phosphate dehydrogenase. Proc. Natl. Acad. Sci. U.S. *57*, 835.

YOSHIDA, A., STAMATOYANNOPOULOS, G. and MOTULSKY, A. G. (1967a). Negro variant of glucose-6-phosphate dehydrogenase deficiency (A^-) in man. Science *155*, 97.

YOSHIDA, A., STEINMANN, L. and HARBART, P. (1967b). *In vitro* hybridization of normal and variant human glucose-6-phosphate dehydrogenase. Nature (Lond.) *216*, 275.

ZELLER, E. A. (1943). Isolierung von Phenyilmichsäure und Phenyltraubensäure aus Harn bei Imbecillitas Phenylpyruvica. Helv. Chim. Acta *26*, 1614.

ZIDERMAN, D., GOMPERTZ, S., SMITH, Z. G. and WATKINS, W. M. (1967). Glycosyl transferases in mammalian gastric mucosal linings. Biochem. Biophys. Res. Commun. *29*, 56.

ZINKHAM, W. H. (1968). Lactate dehydrogenase isozymes of testis and sperm: biological and biochemical properties and genetic control. Ann. N.Y. Acad. Sci. *151*, 598.

ZINKHAM, W. H., LENHARD, R. E. JR. and CHILDS, B. (1958). A deficiency of glucose-6-phosphate dehydrogenase activity in erythrocytes of patients with favism. Bull. Johns Hopk. Hosp. *102*, 169.

Subject index