ADVANCES IN PHYSIOLOGICAL SCIENCES

Volume 6

Genetics, Structure and Function of Blood Cells

Editors
S. R. HOLLÁN
G. GÁRDOS
B. SARKADI

PERGAMON PRESS
AKADÉMIAI KIADÓ

ADVANCES IN
PHYSIOLOGICAL SCIENCES

Proceedings of the 28th International Congress of Physiological Sciences
Budapest 1980

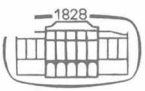
Volume 6

Genetics, Structure and Function of Blood Cells

Editors

S. R. Hollán
G. Gárdos
B. Sarkadi

Budapest, Hungary



PERGAMON PRESS

AKADÉMIAI KIADÓ

Pergamon Press is the sole distributor for all countries, with the exception of the socialist countries.

HUNGARY	Akadémiai Kiadó, Budapest, Alkotmány u. 21. 1054 Hungary
U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada, Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Hammerweg 6, Federal Republic of Germany

British Library Cataloguing in Publication Data

International Congress of Physiological Sciences
(28th : 1980 : Budapest)
Advances in physiological sciences
Vol. 6: Genetics, structure and function of blood cells
1. Physiology - Congresses
I. Title II. Hollán. S. R.
III. Gárdos G. IV. Sarkadi B.
591.1 QP1 80-41876

Pergamon Press	ISBN 0 08 026407 7 (Series) ISBN 0 08 026818 8 (Volume)
Akadémiai Kiadó	ISBN 963 05 2691 3 (Series) ISBN 963 05 2732 4 (Volume)

In order to make this volume available as economically and as rapidly as possible the authors' typescripts have been reproduced in their original forms. This method unfortunately has its typographical limitations but it is hoped that they in no way distract the reader.

Printed in Hungary

FOREWORD

This volume is one of the series published by Akadémiai Kiadó, the Publishing House of the Hungarian Academy of Sciences in coedition with Pergamon Press, containing the proceedings of the symposia of the 28th International Congress of Physiology held in Budapest between 13 and 19 July, 1980. In view of the diversity of the material and the "taxonomic" difficulties encountered whenever an attempt is made to put the various subdisciplines and major themes of modern physiology into the semblance of some systematic order, the organizers of the Congress had to settle for 14 sections and for 127 symposia, with a considerable number of free communications presented either orally or as posters.

The Congress could boast of an unusually bright galaxy of top names among the invited lecturers and participants and, naturally, the ideal would have been to include all the invited lectures and symposia papers into the volumes. We are most grateful for all the material received and truly regret that a fraction of the manuscripts were not submitted in time. We were forced to set rigid deadlines, and top priority was given to speedy publication even at the price of sacrifices and compromises. It will be for the readers to judge whether or not such an editorial policy is justifiable, for we strongly believe that the value of congress proceedings declines proportionally with the gap between the time of the meeting and the date of publication. For the same reason, instead of giving exact transcriptions of the discussions, we had to rely on the introductions of the Symposia Chairmen who knew the material beforehand and on their concluding remarks summing up the highlights of the discussions.

Evidently, such publications cannot and should not be compared with papers that have gone through the ordinary scrupulous editorial process of the international periodicals with their strict reviewing policy and high rejection rates or suggestions for major changes. However, it may be refreshing to read these more spontaneous presentations written without having to watch the "shibboleths" of the scientific establishment.

September 1, 1980

J. Szentágothai

President of the
Hungarian Academy of Sciences

CONTENTS

Regulation of differentiation of haemoglobin structure and function

The role of calcium in red cell membrane transport processes

Structure and function of blood cell membranes

GENETIC REGULATION OF HAEMOGLOBIN SYNTHESIS

Adv. Physiol. Sci. Vol. 6. Genetics, Structure and Function of Blood Cells
S. R. Hollán, G. Gárdos, B. Sarkadi (eds)

GENETIC REGULATION OF HAEMOGLOBIN SYNTHESIS

Susan R. Hollán
National Institute of Haematology and Blood Transfusion, Budapest, Hungary

Human haemoglobin has been the favourite model for the study of a number of basic physiological problems. It has been particularly rewarding in elucidating the relationship between structure and specific functions of biological macromolecules. The genetic polymorphism of haemoglobin provided the possibility for the definition of structural genes and the biochemical evidence for the evolution of a protein family. The classical genetic approach depends on the availability of appropriate mutants. This was amply provided by the phenotipically well characterized abnormal haemoglobins and thalassaemias /syndromes characterized by unbalanced globin chain synthesis/. The genetic lesion giving rise to these deviant phenotypes is derived from studies of their inheritance. Difficulties arise when the role of non-coding, functionally undefined segments of a genome are to be studied. No appropriate mutant is available in this case, since one does not know what property to select or screen for. This difficulty was overcome by the virtue of the new gene splicing and hybridization methods. New rapid DNA sequencing techniques and the relatively large amounts of pure DNA yield by the method of cloning recombinant DNA in bacteria have been invaluable. The most astonishing discoveries that have emerged from DNA sequencing include: selfish genes, overlapping genes, split genes, degenerate genes, the unsuspected features of the mitochondrial genome, including a genetic code that break the rules of the code being universal. The application of the most recent methods in molecular genetics gave rise to the so-called "reversed genetics", wherein nucleic acids are studied or even modified at predetermined positions *in vitro*, and the effects of this intervention on the phenotype of the coded protein are investigated *in vitro* and *in vivo* /Weissman et al., 1977; Müller et al., 1978/. And once again the haemoglobin family proved to be the best working horse for investigations by these emerging techniques. These new approaches provided

a "quantum jump" in our understanding of the regulation of haemoglobin synthesis at the gene level. A brief review cannot include even the most important results in this evergrowing field of science. Without any attempt to be comprehensive the following subjects will be considered: 1. mapping of human globin genes, 2. genetic control of functional units, 3. "selfish" DNA, 4. "parsimonious" DNA, and 5. the significance of the -globin gene duplication.

1. MAPPING OF HUMAN GLOBIN GENES

All normal human haemoglobins consist of two different pairs of globin chains and each of the four polypeptide chains is associated with one haem molecule. Genes coding the globin-chains can be grouped as α-and non-α /viz. ζ, γ, β and δ genes/. The ζ-genes are expressed physiologically during embryonic life, the γ-genes during foetal life, while the genes coding for the major β-and the minor δ-globins are active during adult life. Alpha--genes are expressed during the whole life cycle of the individuum, because the ζ-gene which is active during the early embryonic life can be regarded as a primitive α-chain gene /Kamuzora et al., 1974/ and α-chains also can be detected already in early embryos /Figure 1./.

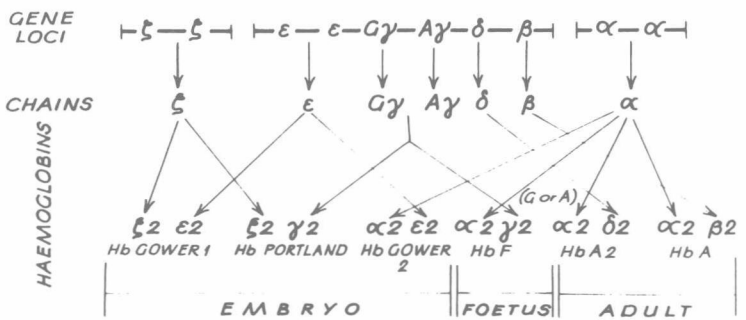


Fig. 1. Human haemoglobins at different stage of development.

In mammals, including man, the genes coding the structure of globin chains are represented but a few times in the cellular DNA /Nienhuis and Benz, 1977/. This comprises only one part of DNA in several million. This made it impossible to isolate and examine a single or a few copies of genes, to define the number and the linkage relationships of specific genes by indirect methods. The discovery of DNA restriction enzymes and the viral reverse transcriptase rendered this problem amenable to attack.

DNA restriction endonucleases are bacterial enzymes capable of splicing DNA at specific base sequence sites /Nathans and Smith, 1975/. Bacteria most probably synthetize these enzymes to protect them from infecting viruses. Their own DNA is protected from being chopped up by their own restriction endonucleases by methylation of the specific DNA sites susceptible to the attack of these enzymes. A continuously growing number of restriction endonucleases are now available. Each enzyme cuts DNA at different and highly specific base sequences.

When high molecular weight DNA is being digested with sequence specific restriction enzymes many fragments are generated. These fragments can be separated by electrophoresis in agarose on the basis of their size /Southern, 1975/. The location of specific globin sequences among the DNA fragments can be determined by molecular hybridization with radioactive probes [the so-called complementary DNAs /cDNAs/] and autoradiography either after transfer of the fragments to filter paper /"blotting"/ or in situ/Shinnick et al., 1975/. The radioactively labeled DNA copies are produced by an enzyme, the reverse transcriptase on messenger RNA templates for specific globins in the presence of all building blocks required to make DNA. The resulting cDNA has a base composition precisely complementary to the mRNA from which it was copied. If the total cellular /double stranded/ DNA is heated, the two DNA strands will come apart. The radioactive DNA for a specific globin gene is added to the mixture of single stranded DNAs, which thereafter is allowed by cooling to reanneal again to double stranded DNA. The cDNA will anneal /hybridize/ only if it finds a complemetary sequence /a globin gene/ in the total cellular DNA.

To permit rational selection of enzymes for digestion of cellular DNAs the action of several enzymes on radioactive, in vitro synthetized double stranded DNAs copied from globin mRNAs /cDNAs/ by transcriptase had to be examined initially /Orkin, 1978/. Knowledge of restriction enzyme cleavage sites within human globin gene sequences enabled the design and execution of detailed mapping of haemoglobin genes.

The human α, β, γ and δ globin genes are by now completely mapped by cutting the globin genes either within or outside the coding sequences /Orkin, 1978; Flavell et al. 1978; Mears et al., 1978; Little et al., 1979/. Direct determination of gene linkage and intergene distances has been possible.

By the use of the recombinant DNA technique restricted fragments of DNA, containing globin genes, can be inserted into the DNA of plasmids or bacteriophages in certain bacteria. Under specific conditions large amounts of pure globin genes can be generated in these "bacterial

factories". This is the only means to obtain adequate amounts of pure genes to sequence their primary structure. The rDNA technique renders extremely pure probes for locating globin genes in the course of different investigations.

One of the most surprising findings of gene mapping was the detection that globin genes /as well as all so far tested eukaryotic genes/ are discontinuous /Williamson, 1977; Leder, 1978/: they contain one or more non-coding inserts so-called introns within the structural gene.

Genetic analysis of appropriate families with β - and δ -chain mutants strongly supports the presence of a single locus for each of these two globins. The locus controlling the synthesis of δ -chains can, however, be regarded as a duplication of the β -chain locus leading to the formation of the minor adult component HbA_2 /$\alpha_2\delta_2$/. Schroeder et al. /1968/ presented evidence for multiple structural genes coding the γ -chain of human foetal haemoglobin /$\alpha_2\gamma_2$/. The two γ-chains differ only in the nature of the residue at position 136 /Gγ = γ^{136} Gly and Aγ = γ^{136} Ala. A third type of human γ -globin chain /T$^{D}\gamma^{75}$ Thr / has been described in which isoleucine is replaced by threonine at position 75. It was suggested, however, that T γ results from polymorphism in the sequence of the A γ globin rather than being the gene product of a third distinct γ -chain locus /Schroeder et al., 1979/.

The results of recent cell hybridization experiments suggest that the non-α chain loci are linked together on the short arm of chromosome 11 /Deisseroth et al., 1978/, comfirming the results of previous genetic studies on $\delta\beta$ /Hb Lepore/ and $\gamma\beta$ /Hb Kenya/ cross-over variants. The embryonic ε -globin gene is located on the 3' side of the $^{G}\gamma$. Pseudo-β-like genes, which are not being expressed, were also found on the short arm of chromosome 11. The human α -globin structural gene had been located on chromosome 16 /Deisseroth et al., 1977/. The duplication of the human α -chain locus will be discussed later. The duplicated embryonic ζ -globin genes are located 5' to the α-globin gene. Pseudo-α -like genes lie between ζ and α genes. /Fig. 2./

Fig. 2. Human haemoglobin genes.

2. GENETIC CONTROL OF FUNCTIONAL UNITS

The basic function of haemoglobin is the transport of oxygen. Haemoglobin performs this function most effectively by binding oxygen cooperatively and adapting its overall affinity appropriately to physiological conditions. Regulation of oxygen affinity is achieved by the preferential binding of protons, carbon dioxide and organic phosphates to deoxyhaemoglobin /to the low affinity, tense /T/ state/. The relaxed /R/ quaternary structure of oxyhaemoglobin represents the high affinity state for oxygen binding.

Investigations on structure-function relationships in normal oxy- and deoxyhaemoglobin, as well as in abnormal haemoglobins with altered function, - pioneered by Perutz and coworkers - revealed five functionally important structural domains in the haemoglobin molecule:

/a/ Residues in contact with the haem

These provide the hydrophobic environment necessary for the reversible binding of oxygen and also transmit structural changes triggered by oxygenation of the haem iron, to the $\alpha 1\beta 2$ subunit interface.

/b/ Residues in the $\alpha_1\beta_2$ contact

The bonding of subunits at the $\alpha_1\beta_2$ interface is altered during the switch in quaternary structure but the same residues remain involved in both R and T state.

/c/ Residues in the $\alpha_1\beta_2$ contact remain unaltered by the quaternary structural changes of oxygenation and deoxygenation.

/d/ Residues of the β-chain that bind 2,3-DPG, the most potent allosteric effector of haemoglobin.

/e/ Residues involved in the Bohr-effect

Recent results revealed important relationship between protein functional units and the mosaic structure of eucaryotic genes. Gilbert /1978/ raised a new hypothesis suggesting that the coding sequences /exons/ of genes correspond to functional parts of the proteins, while the non-coding introns speed evolution by allowing new proteins to be constructed from parts of old ones. Specific recombinations between introns can bring together exons into a transcription unit to make special differentiation products. Sakano et al. /1979/ have confirmed this hypothesis by showing that each of the six known functional units in the immunglobulin G heavy chains is coded by a separate exon. In the same line Blake /1979/ and Gilbert /1979/ pointed out that the central exon of the β-globin gene corresponds very closely to the protein fragment that defines the haem pocket. Eaton /1980/ tested the hypothesis of Gilbert by considering the relationship between the coding sequences of α-and β-globin genes and the corresponding parts of the complete, tetrameric haemoglobin molecule. It is concluded that in addition to most of the haem contact residues most of the α- and β-chain residues in the $\alpha_1\beta_2$ contact are coded by the central sequence in the α- and in the β-globin genes. Most of the residues involved in the $\alpha_1\beta_1$ contact are coded by the third sequence of both, the α- and the β-genes. The distribution of residues involved in binding allosteric regulators /protons, carbon dioxide, and 2,3-DPG/ indicates no marked correlation, although the most important residues are contained in the first and third gene sequences. The close relationship between functional properties of the three polypeptide sequences of haemoglobin and the corresponding three coding sequences raises the possibility that the exon/intron mosaic pattern of the eucaryotic

gene structure may carry the frozen evolutionary history of proteins.

The primitive Hb is assumed to be a monomer. Eaton /1980/ argues that following gene duplication and point mutations in the central coding sequence of the primitive α- and β -globin genes led to the development of the $\alpha_1\beta_2$ dimer that was capable of cooperative binding. The monomer lamprey haemoglobin aggregates in the deoxy state to form a homodimer /very probably of the $\alpha_1\beta_2$ type/ achieving thereby a maximum Hill n value of 2. An increase in cooperativity toward the current n value of 2.8 to 3.0 took place with the evolution of the $\alpha_1\beta_1$ contact to form a tetramer. The emergence of the Bohr effect and DPG binding residues are more recent events of evolution.

3. "SELFISH" DNA

The estimated number of genes in the human genome appears too few to account for the 3×10^9 base pairs found for haploid set of DNA. It has been suggested repeatedly /Ohno, 1972; Hinegardner et al., 1976; Galan et al., 1976; Cavalier-Smith, 1978/ that the DNA of higher organisms consists of a minority of sequences with highly specific functions and a majority of sequences /including some of the so-called single copy DNA/ with little or no specific functions. In other words, they do not code for protein since they do not occur at all in messenger RNA /mRNA/. Orgel and Crick /1980/ share the opinion with many others that it does not seem very plausible that all these extra DNA is needed for gene control, although some parts of it must be. The conviction has been growing that this extra DNA conveys little or no selective advantage to the organism and consequently has been considered as "junk".

The catch phrase "selfish gene" has been coined by Dawkins /1976/. Orgel and Crick /1980/include in addition to repetitive DNA the non-coding intervening sequences /introns/ of genes, and parts of the DNA sequences between genes into the category of selfish DNA and compare the spread of these sequences which do not contribute to the phenotype of the organism to the spread of a not-too-harmful parasite /a useless burden to the cell that contains it/ and even consider it as a "cancer of the genome".

The primary transcript of DNA, the so-called heteronuclear RNA /hn RNA/ contains still large amounts of extra nucleic acid. Recent gene mapping and direct nucleotide sequence analysis of clonal genes for globin /and other proteins/ have revealed that the introns are initially transcribed into Hn RNA. They are removed /spliced/ by an enzymatic process in the cell nucleus preceding the transport of mRNA to the cell cytoplasm.