CURRENT TOPICS IN HEMATOLOGY

VOLUME 5

Editors

Sergio Piomelli

Stanley Yachnin

CURRENT TOPICS IN HEMATOLOGY

VOLUME 5

Editors

Sergio Piomelli

Division of Pediatric Hematology-Oncology
Columbia-Presbyterian Medical Center
New York, New York

Stanley Yachnin

Department of Medicine
University of Chicago
Chicago, Illinois

Alan R. Liss, Inc., New York

Address all inquiries to the Publisher
Alan R. Liss, Inc., 41 East 11th Street, New York, NY 10003

Printed in the United States of America.

LC 78-19681
ISBN 0-8451-0354-7
ISSN 0190-1486

Contents of Previous Volumes

Contents of Volume 1

Contents of Volume 2

Contents of Volume 3

Contents of Volume 4

Contents

Contributors

Arthur Bank, Departments of Medicine and of Human Genetics and Development, Hammer Health Sciences Center, College of Physicians and Surgeons, Columbia University, New York, NY 10032 **[1]**

Danny Tsun-Yee Chiu, Bruce Lyon Memorial Research Laboratory, Children's Hospital Medical Center, Oakland, CA 94609 **[63]**

A. Jacobs, Department of Haematology, Welsh National School of Medicine, Cardiff, UK **[25]**

Margaret Karpatkin, Departments of Pediatrics and Medicine, New York University School of Medicine, New York, NY 10016 **[113]**

Simon Karpatkin, Departments of Pediatrics and Medicine, New York University School of Medicine, New York, NY 10016 **[113]**

Bertram Lubin, Bruce Lyon Memorial Research Laboratory, Children's Hospital Medical Center, Oakland, CA 94609 **[63]**

Robert S. Schwartz, Bruce Lyon Memorial Research Laboratory, Children's Hospital Medical Center, Oakland, CA 94609 **[63]**

Ugo Testa, Unité de Recherche en Génétique Moléculaire et en Hématologie, INSERM U.91, Hôpital Henri Mondor, 94010 Creteil, France **[127]**

The number in brackets is the opening page number of the contributor's article.

Current Topics in Hematology 5:1–23

Genetic Defects in the Thalassemias

Arthur Bank

Departments of Medicine and of Human Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, New York 10032

I. NORMAL GLOBIN BIOSYNTHESIS

Human hemoglobins are composed of tetramers of globin chains and four heme molecules [1–3]. Two different globin chains are present in each of the human hemoglobins—normal adult hemoglobin is composed of two α- and two β-globin subunits. Hemoglobin A_2, a minor constituent, contains two α- and two δ-globin chains (Fig. 1). Fetal hemoglobin, the major hemoglobin type present during fetal life, is composed of two α- and two γ-chains. Structural analysis at the globin level indicates that α- and non-α-globin genes diverged at an early stage in evolution. The γ- and β-globin chains are

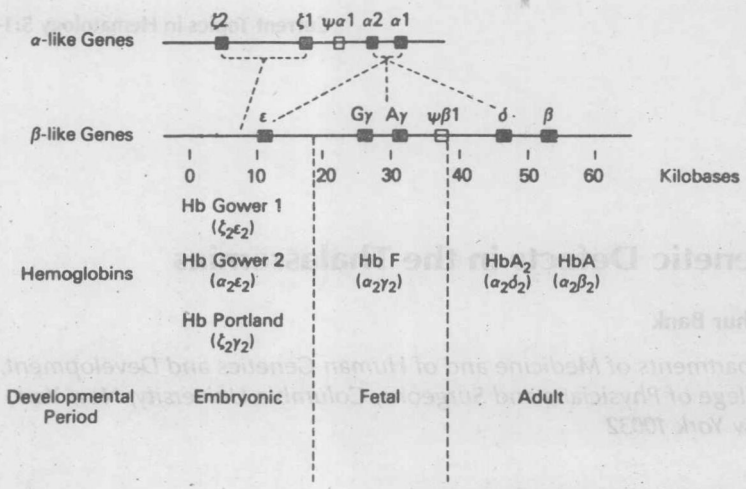

Fig. 1. Human hemoglobins. The α-like and β-like genes are shown. The embryonic, fetal, and adult hemoglobins are also indicated. The distances between the β-like genes are in kilobases.

much more similar in structure than the α- and β-chains. The δ- and β-chains differ by only 10 of 141 amino acids.

In addition to adult and fetal hemoglobins, there are several embryonic hemoglobins which are present during embryonic and fetal life (Fig. 1). These include the Gower hemoglobins, composed of either ε- and ζ-embryonic globin chains (Gower 1), or ε-and α-chains (Gower 2).

The details of the biosynthesis of all of the globin chains are now available by the cloning and sequencing of each of the human globin genes. The results have shown that globin biosynthesis is more complex than was previously thought. Intervening sequences (IVS) are present in the globin genes and are transcribed into nuclear RNA; although the function of these sequences is not yet clear, they have been preserved through evolution at specific locations within each of the globin genes, and presumably have functional significance.

A. Globin Genes

The structure and organization of all of the functional human globin genes are now known [4, 5] (Figs. 1, 2). Initial studies using restriction analysis of cellular DNA led to the conclusion that two γ-, one δ-, and one β-globin genes were present on a single stretch of DNA [1, 6–9]. These experiments showed that the relative distance between the γ-genes was approximately 2.5

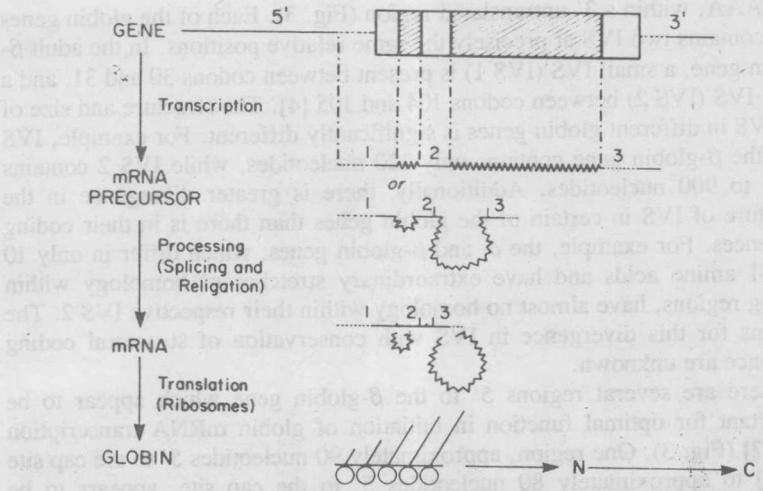

Fig. 2. Globin biosynthesis. The globin gene with its 5′ and 3′ flanking sequences is shown. The clear areas of the gene are the coding sequences and the hatched areas the IVS. Coding regions 1, 2, and 3 are transcribed into RNA together with the IVS (sawtoothed segment). The IVS are then removed by splicing, and the coding regions of RNA are religated to form mature globin mRNA. The lowest line shows globin mRNA being translated into globin on ribosomes (circles).

kilobases (kb), the distance between the γ- and δ-genes was approximately 15 kb, and between the δ- and β-genes approximately 5.5 kb (Fig. 1). Restriction analysis also indicated that each of the γ-, δ-, and β-genes contained IVS. The presence of two linked α-globin genes was also first shown by restriction analysis [10].

Definitive analysis of the structure and organization of the human globin genes resulted from the cloning and nucleotide sequencing of each of these genes and their flanking sequences [11–16]. Clones containing each of these genes have been isolated, and overlapping clones have been identified which permit the organization of the ϵ-, γ-, δ-, and β-genes with respect to each other [4] (Fig. 1). Similarly, overlapping clones have indicated the organization of the α-globin genes [4]. Here, two structural α-genes are located 3′ to a pseudo-α-gene which in turn is located 3′ to two embryonic ζ-genes, one of which is active and one of which is inactive (Fig. 1).

Each of the globin genes contains a "cap" site approximately 50–60 nucleotides 5′ to the initiating codon, has approximately 140–146 coding sequences (codons) within it, and has a polyadenylation (poly-A) signal,

AATAAA, within a 3′ untranslated region (Fig. 3). Each of the globin genes also contains two IVS at precisely the same relative positions. In the adult β-globin gene, a small IVS (IVS 1) is present between codons 30 and 31, and a large IVS (IVS 2) between codons 104 and 105 [4]. The structure and size of the IVS in different globin genes is significantly different. For example, IVS 1 of the β-globin gene contains only 120 nucleotides, while IVS 2 contains close to 900 nucleotides. Additionally, there is greater divergence in the structure of IVS in certain of the globin genes than there is in their coding sequences. For example, the δ- and β-globin genes, which differ in only 10 of 141 amino acids and have extraordinary stretches of homology within coding regions, have almost no homology within their respective IVS 2. The reasons for this divergence in IVS with conservation of structural coding sequence are unknown.

There are several regions 5′ to the β-globin gene which appear to be important for optimal function in initiation of globin mRNA transcription [15–17] (Fig. 3). One region, approximately 90 nucleotides 5′ to the cap site (−90) to approximately 80 nucleotides 5′ to the cap site, appears to be important for optimal transcription as shown by studies of the rabbit globin gene [17]. Similarly, another region approximately 50–60 nucleotides 5′ to the cap site (−50 and −60), the so-called CAAT box, appears to be important in optimal function. This region is so named since the nucleotide sequence, CAAT, is usually present in this region. A third region, the so-called TATA or ATA box (the Goldberg-Hogness box), is at −30 to −40 5′ to the β-gene, and is again required for optimal transcription of this gene. These three 5′ flanking sequences are all considered to be "promoter" sequences, important in the regulation of transcription of the globin genes. Single base changes in each of these regions are associated with decreased globin mRNA transcription. There is significant heterogeneity in the sequences 5′ to the

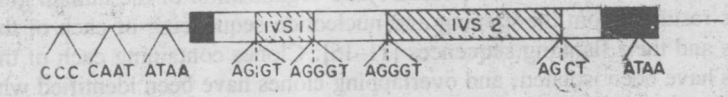

Fig. 3. Key nucleotides within and surrounding globin genes. The critical nucleotides within and surrounding the β-globin gene are shown. In the region 5′ to the gene (single horizontal line to the left), three critical regions are present: A CCC sequence 86–88 nucleotides 5′ to the gene, a CAAT sequence, and an ATAA sequence. A single base change in the middle C of the CCC sequence leads to β^+-thalassemia (Table I). Changes in the ATAA sequence also lead to β^+-thalassemia (Table I). The black areas show the 5′ and 3′ untranslated regions. The sequence responsible for polyadenylation, ATAA, is in the 3′ untranslated region. The conserved nucleotides at the 5′ and 3′ ends of IVS 1 and IVS 2 are shown, and single changes can cause β^0-thalassemia (Table I).

cap site in different globin genes; these sequences differ in the β-, δ-, γ-, ϵ-, ζ-, and α- genes.

Nucleotide sequences within and surrounding the IVS of different globin genes show remarkable homology and conservation at their 5′ and 3′ ends [18, 19]. At the 5′ end of the IVS of all of the globin genes, as in most other genes studied to date, there is an invariant GT dinucleotide. In many cases, additional A and G residues adjacent to this GT dinucleotide are present, and the sequence AGGT provides a "stronger consensus" sequence at the 5′ splice junction [19]. An invariant AG dinucleotide is present at the 3′ end of all IVS in the globin genes. This dinucleotide is usually preceded by 12–15 pyrimidines which appear to strengthen this site as the 3′ splice junction. As will be discussed later, changes in a single nucleotide of the invariant dinucleotides, either 5′ or 3′ extremities of IVS, lead to dramatic alterations in the ability to form mature globin mRNAs.

Transcription of the human globin genes begins at the cap site and includes the three coding regions (exons) and the two IVS past the 3′ untranslated region of these genes, and ends 3′ at the so-called poly-A "addition" site (Fig. 2). Studies both in intact cells and in cell-free systems show that there is modification of the cap by unusual methylated bases at the 5′ end, and poly-A addition at the 3′ end of these primary RNA transcripts soon after transcription.

In addition to a 5′ cap structure and 3′ poly-A addition, the formation of mature human globin mRNA requires the accurate removal of IVS present in the primary transcript. This is presumed to be an enzymatic process generally known as RNA processing, and in this process, the IVS are accurately "spliced out" (Fig. 2). There is some overlap in the nucleotide sequences at the 5′ and 3′ ends of IVS, which allows some ambiguity in the precise nucleotides spliced out, while still ensuring retention of the proper coding sequence continuity required in the exons. The role of secondary structure within or surrounding IVS, and the precise participation of enzymes, other proteins, or the RNA itself in the process of splicing, is completely unknown, and represents an area of active investigation. The order in which IVS are spliced out and whether this order is of significance is also not clear, although it is obvious that both IVS must be completely spliced out to produce mature messenger RNA.

Several methods are available for analyzing the RNA produced either in intact cells or by globin genes transferred into cells [20–23]. Radioactively labelled DNA probes can be used to define RNA species present. The hybrid formed between the radioactively cloned DNA and the RNA can be analyzed using the enzyme S1. This enzyme cleaves single-stranded DNA or RNA,

but not double-stranded nucleic acids. The size of the protected labeled DNA fragments defines the regions represented in RNA. The size of RNA precursors or mature RNA products can be identified accurately by so-called Northern blot analysis [23]. In this analysis, agarose gels containing the RNA are used to separate RNA components on the basis of size. The RNA is transferred from gels to nitrocellulose filters, and the filters hybridized to specific DNA probes which identify globin gene-specific fragments. This methodology can identify both the size and the amount of RNAs in cells. Newly transcribed RNAs can be identified by radioactively labeling the newly synthesized RNA and using unlabeled cloned DNA probes for their analysis [20].

B. Globin mRNA Translation

Globin mRNA precursors are spliced in the nucleus to produce mature globin mRNA. This mature globin mRNA is then transported to the cytoplasm. In the cytoplasm, the mRNA becomes associated with polyribosomes. In the presence of appropriate tRNAs, amino acids, and enzymes, globin chains are synthesized on the polyribosome-mRNA template. A variety of protein factors required for translation have recently been identified, and are necessary for normal initiation, elongation, and termination of translation of globin chains in the cytoplasm. Individual α- and β-globin mRNAs are associated with different-sized polysomes, and the α- and β-globin chains produced initially combine to form $\alpha\beta$-dimers which rapidly acquire heme groups. The $\alpha\beta$-dimers then associate to form stable hemoglobin tetramers. Individual α- and β-globins are not as stable as the mixed normal $\alpha_2\beta_2$-hemoglobin tetramers. Excess α-globin chains tend to aggregate and become insoluble and in high enough concentrations will precipitate. β-Globin subunits unassociated with α-globin subunits combine to form a β_4-tetramer known as hemoglobin H (Hgb H), which is less stable than normal hemoglobin.

II. THE β-THALASSEMIAS

A. Clinical Variants

The β-thalassemias are due to decreased or absent β-globin biosynthesis leading to decreased or absent amounts of hemoglobin A (Hgb A) [1–3]. Heterozygotes for β-thalassemia inherit a defective β-globin gene from only one parent, while homozygotes inherit defective β-globin genes from both parents. Although there are many different β-thalassemia genes, there are basically two types: those associated with the production of some β-globin,

called β^+-thalassemia genes, and those which produce no β-globin, called β°-thalassemia genes. It is now known from a variety of studies that individuals homozygous for β-thalassemia can have two β^+- or two β°-thalassemia genes, or one β^+-and one β°-thalassemia genes. When two β°-thalassemia genes are present, the patient has β°-thalassemia, and no β-globin is produced. When at least one β^+-thalassemia gene is present, then the individual continues to produce small amounts of β-globin, and has the β^+-thalassemia phenotype. Many individuals have now been identified who are heterozygous for both β^+- and β°-thalassemia. Different β^+-thalassemia genes have different outputs of β-globin mRNA and, consequently, of β-globin. For example, it is now known that certain β^+-thalassemia genes in blacks are associated with significantly more β-globin production than other β^+-thalassemia genes present in Mediterranean individuals and those of other ethnic groups. Thus, black individuals with homozygous β^+-thalassemia are usually less severely affected than those in other ethnic groups. Although studies, to date, are limited, there appears to be no significant correlation between the severity of β^+-thalassemia in Mediterranean individuals who are homozygous for two β^+-thalassemia genes, and those heterozygous for both one β^+-and one β°-thalassemia genes. This presumably is due to the fact that all or most of the β^+-thalassemia genes in individuals of Mediterranean extraction manifest little β-globin synthesis, and not enough Hgb A is produced by most variant β^+-thalassemia genes to significantly alter the clinical severity in these different individuals.

While the fundamental defect in β-thalassemia is due to decreased or absent β-globin synthesis, the pathogenesis of β-thalassemia is largely due to the continued normal production of α-globin chains [1–3]. The large amounts of α-globin chains produced in the bone marrow have no significant amounts of non-α-chains with which to combine to form stable hemoglobins. Because of this situation α-globin chains aggregate and precipitate in cells in the earliest erythroid precursors in the bone marrow. These α-chains disrupt the normal metabolism of nucleated red cells and cause their premature destruction. Thus, although the production of erythroid cells in the bone marrow in β-thalaessemia is remarkably increased, the number of functional erythroid precursors which mature normally and produce hemoglobin is markedly diminished. In addition, cells that do produce enough Hgb A and Hgb F to survive and reach the peripheral blood also contain excess α-globin. This excess α-globin material in the circulating erythroid cells causes increased destruction of these cells by the spleen and other phagocytic organs.

Different patients with β-thalassemia also vary in their fetal hemoglobin production. The production of γ-globin chains and fetal hemoglobin has two

effects. First, increased γ-chains leads to increased Hgb F, producing more viable oxygen-carrying red cells. Second, the increased γ-globin production leads to less accumulation of excess α-globin chains, decreasing the severity of the thalassemia. There is variation in the extent of hemoglobin F production in different patients with β-thalassemia in different ethnic groups, which appears to be largely independent of the type of β-thalassemia genes present in these patients. The precise relationship between the genetic factors regulating fetal hemoglobin production and those involved in altering β-globin production in thalassemia has not yet been defined.

B. Globin Biosynthesis

Globin biosynthesis in thalassemia is measured by using a radioactive amino acid to quantify the relative amounts of α- and β-globin chains produced in reticulocytes in peripheral blood, nucleated erythroid cells, and reticulocytes in bone marrow. Many studies over the past two decades have revealed that, while normally there is relatively balanced α- and β-globin synthesis in both nucleated red cells and reticulocytes, in most cases of homozygous β-thalassemia, there is a decrease in the amount of β-globin synthesis as compared to α-globin synthesis. In heterozygous β-thalassemia, there is also a reduced relative amount of β-globin synthesized as compared to α-globin in peripheral blood reticulocytes. By contrast, in bone marrow cells, there are relatively equal amounts of synthesized α- and β-globins in heterozygotes.

C. Globin mRNA

Several studies have determined that in β^+-thalassemia there is a reduction in the amount of normal β-globin mRNA. Two different methods have been used to determine β-globin mRNA. Globin mRNA has been isolated from reticulocytes and bone marrow cells and translated in vitro using cell-free systems. Using a variety of cell-free systems, ranging from wheat germ to rabbit retculocytes, there is a decreased amount of β-globin as compared to α-globin biosynthesis using mRNA from β-thalassemia homozygotes, and to a lesser extent from β-thalassemia heterozygotes as well. RNA from individuals homozygous for β°-thalassemia produce no β-globin. To determine whether β-globin mRNA in β-thalassemia is due to a decreased amount of normal β-globin mRNA or to a defective β-globin mRNA, complementary DNA (cDNA) was synthesized which hybridized specifically to β-globin mRNA, and was used to quantify the amount of β-globin mRNA in different individuals with β-thalassemia. In individuals with β^+-thalassemia, it was shown, using cDNA hybridization, that there was a 3–10-fold decrease in the

amount of β-globin mRNA sequences present in the reticulocytes and bone marrow precursors of these patients. In β°-thalassemia, there was heterogeneity in the results obtained by cDNA hybridization. In some individuals with β°-thalassemia, there were markedly diminished amounts of β-globin mRNA consistent with the absence of β-globin mRNA sequence. By contrast, in other individuals with β°-thalassemia, there were somewhat diminished complements of β-globin mRNA sequences, but β-globin mRNA sequences were present. These latter results indicated that in many individuals with β°-thalassemia, the defects were due to abnormal β-globin mRNAs which resulted in the absence of β-globin production. These preliminary assumptions, based on analyses of β-globin mRNA by cDNA hybridizations, have been subsequently borne out by isolation of abnormal β-globin genes in β°-thalassemia, and characterization of the abnormal β-globin mRNAs produced by these genes.

D. Gene Defects in the Thalassemias

1. Restriction enzyme analysis. Two types of analyses—restriction enzyme studies [6–10] and globin gene cloning [11–14]—have led to the detection of more than 30 different types of gene defects in the β-thalassemias. Restriction enzyme analysis involves the isolation of high molecular weight DNA from peripheral blood white cells, or other of the patient's tissue material. Cleavage of the DNA is then performed using one of many available restriction enzymes. These restriction enzymes cut specific nucleotide sequences, and, thus, lead to an ordered array of DNA fragments. The restricted cellular DNA fragments are separated by agarose gel electrophoresis, and then are transferred to nitrocellulose filters. The particular DNA fragments which contain the β-globin genes are identified by the use of a highly radioactive DNA probe prepared by using either cDNA or cloned DNA fragments that have been radioactively labeled. The filters are hybridized to the radioactively labeled DNA, and after washing are subjected to radioautography. Specific bands of DNA representing the β-globin genes are seen depending on the particular restriction enzyme used and the specific fragment of DNA utilized as probe.

Restriction enzyme analysis has been used to detect a few thalassemia mutations: deletions of parts of the β-globin gene and single nucleotide defects. Deletion of the 3′ end of the β-globin gene occurs in a subset of Indian patients with β°-thalassemia [24, 25]. Several different β-thalassemia genes caused by single nucleotide changes have also been detected using restriction enzyme analysis. The first of these involved the use of the enzyme Hph I. which recognizes the nucleotide sequence present at the 5′ end of β-