REVIEW

Advances in Thalassemia Research

By Arthur W. Nienhuis, Nicholas P. Anagnou, and Timothy J. Ley

The thalassemias are hereditary hemolytic anemias characterized by decreased or absent synthesis of one of the globin subunits of the hemoglobin molecule. In the α-thalassemias, decreased synthesis of α-globin results in accelerated red cell destruction because of the formation of insoluble HbH (β₄) inclusions in mature red cells. The greater clinical severity of the β-thalassemia reflects the extreme insolubility of α-globin, which is present in relative excess because of decreased β-globin synthesis. Insoluble α-globin precipitates in developing erythroblasts, leading to marked ineffective erythropoiesis. Severe disease does occur among the α-thalassemias, but only when no α-globin is produced. The thalassemias are among the most common genetic diseases of man. The high gene frequency results from a selective advantage of the thalassemia phenotype in heterozygotes, where it protects from severe malaria.

A quantitative decrease in globin synthesis without a qualitative change in the gene product must reflect an alteration in gene function or regulation. The recent explosion in our understanding about gene organization and structure has been matched by characterization of several specific thalassemia mutations at the level of DNA sequence. Many well studied examples of mutations that alter specific steps in gene expression, e.g., gene transcription, RNA processing, and mRNA translation, will form the focus for this review.

The current status of several approaches to the treatment of severe thalassemia will also be summarized. Modifications of the transfusion regimen and use of subcutaneous desferrioxamine have resulted in improved conventional treatment, but the problem of ultimately fatal iron overload may not yet be completely solved. Bone marrow transplantation and pharmacologic stimulation of HbF production are now being explored as potentially useful, but still highly experimental approaches, for treatment of severe β-thalassemia. Insertion of normal globin genes into bone marrow stem cells of thalassemic individuals offers the possibility of cure of this disease. Gene therapy is not yet practical, although many successful strategies are being employed to transfer genes into cells in tissue culture; the application of these techniques to the problem of thalassemia in the future seems quite probable.

There have been major advances in the prenatal diagnosis of the thalassemias and hemoglobinopathies during the past several years. These advances are beyond the scope of this review, but will be the subject of detailed consideration in other issues of Blood.

Organization and Structure of Human Globin Genes

Organization

The human globin genes occur in multigene clusters; the α-like genes are found on chromosome 16 and the β-like genes on chromosome 11. There are two fully functional and nearly identical α and γ-genes on each chromosome, reflecting their duplication during evolution and subsequent gene correction events. The closely related δ and β-genes reflect gene duplication, but subsequent sequence “mutations” in the regulatory region of the δ-gene have rendered this gene functionally insignificant. Therefore, there are only two active β-genes in each human diploid erythroid cell. Both the α and β-gene clusters include pseudogenes. These are thought to be byproducts of evolution, representing previously functional genes in which sequence alterations in coding or regulatory regions have led to inactivation. The DNA between the individual globin genes includes copies of many repetitive DNA sequence families, but no other functional genes are known to occur within the α and β-gene clusters.

Structure

A prototypical globin gene (Fig. 2B) includes three coding blocks (exons) separated by two introns or intervening sequences. Conserved sequences important for gene function are found just before the mRNA coding sequences, at the exon-intron boundaries, and at the end of the mRNA coding sequences. The globin gene promoter includes sequences at the 5′ end that are essential for RNA polymerase binding and for accurate and efficient initiation of RNA transcription. Differences among the various β-like globin genes with respect to the arrangement of conserved sequences in the promoter region (Fig. 2A) may be relevant to their
selective expression during development (discussed below). The exon-intron boundary sequences are crucial for the accurate removal of RNA transcribed from the introns and for the subsequent precise splicing of coding sequences to form a functional mRNA molecule (Fig. 2C). Transcription appears to continue beyond the position on the DNA strand that corresponds to the 3' end of RNA coding sequences, conserved sequences found at this position may be involved in cutting the RNA transcript and the addition of adenosines to form the poly A tract. The poly A tract may be important for nuclear to cytoplasmic mRNA transport and for mRNA stability in the cytoplasm.

MOLECULAR MECHANISMS OF THALASSEMIA

The thalassemias must be considered with the realization that these are cis acting mutations. A cis acting...
mutation affects only the gene on the chromosome on which the mutation occurs; a trans acting mutation influences expression of both genes at a chromosomal locus. The output of only one of the two β-genes in erythroid cells is affected by a specific thalassemia mutation. For example, in the cells of individuals who are doubly heterozygous for a sickle (βS) and a thalassemia gene, total βS-globin synthesis is normal or even increased, whereas β-globin synthesis is decreased (β−) or absent (β0). As a general rule, cis acting mutations affect gene structure directly, rather than altering the function of another molecule involved in gene expression. Indeed, all thalassemia genes that have been carefully studied to date have been found to contain mutations that directly alter gene structure and, thereby, gene function. Gene deletion accounts for many, but not all, α-thalassemia mutations. The α-genes are apparently predisposed toward deletion because of tandemly duplicated sequences that occur in this cluster.5,5 Single nucleotide substitutions (or deletion of one or a few nucleotides) in otherwise intact genes account for the vast majority of β-thalassemia mutations. The only exception, other than the extensive deletions discussed below, is a mutation characterized by deletion of the 3′ portion of the β-globin gene and adjacent flanking sequences, which is observed in a few patients of Indian ancestry.19–21

### Promoter Mutations

The promoter of a gene is defined as those DNA sequences that are essential for accurate and efficient initiation of transcription. Promoter sequences are found near the site at which initiation of transcription begins and function in only a single orientation with respect to the coding sequences. Comparison of DNA sequences found just upstream to the coding portions of the α and β-globin genes of several species provided the first clues as to the nature of DNA sequences that function as a promoter.7,14 A conserved block of nucleotides is found at a position just before the site corresponding to the 5′ end of mRNA, and several additional blocks are found within 150 nucleotides upstream. The position of the conserved blocks for the human β-like genes is shown in Fig. 2A. A conserved block of adenine-thymine (AT)-rich sequences, referred to as the TATA box, begins 28–31 basepairs (bp) upstream from the modified methylated 5′ end of mRNA (CAP) site (Fig. 2A). Forty-six to 58 nucleotides upstream is a second conserved block that includes the pentanucleotide segment, CCAAT, leading to the designation of this sequence as the “CAT” box. Significant evolutionary divergence has occurred in the DNA sequences between these conserved blocks. Note that the “CAT” box and 18 nucleotides of surrounding DNA sequence are duplicated in the γ-gene promoter. The β-globin gene promoter also includes tandemly duplicated conserved blocks located 86 and 108 nucleotides upstream (Fig. 2A). This sequence is not found upstream from the δ-gene and is found only once upstream from the γ and ε-genes. The α-globin gene promoter also lacks this conserved block of nucleotides. The striking differences among the sequences of the ε, γ, and β-globin gene promoters invite the hypothesis that promoter structure is important for developmental regulation of globin gene expression.

Simple sequence comparison could not lead to identification of all sequences important for promoter function. Hence, strategies have been devised for removing DNA sequences to create genes that are truncated at various positions in the promoter region. Specific small deletions and single nucleotide substitutions have also been made to create additional globin gene “promoter mutants.” The function of these mutated genes has been compared to that of the normal gene after introduction of recombinant plasmids containing each into monkey kidney or HeLa cells.14,22,23 Globin genes are actively transcribed in such cells; after 48 hr, RNA is harvested for analysis. The relative amount of correctly initiated globin mRNA provides a measure of promoter function.

These gene expression studies have provided a functional profile of the β-globin gene promoter.14,23 The promoter extends upstream from the CAP site for 110 bp (−110). The conserved CCAAT and TATA boxes are indeed important—deletion of these blocks, single nucleotide substitutions within these sequences, or changes in the relative distance between the two cause a marked quantitative decrease in promoter function. The duplicated sequences between −85 and −110 are also important (Fig. 2B)—deletion of these sequences abolishes promoter function, and single nucleotide substitutions can cause a substantial decrease in the amount of correctly initiated β-globin mRNA. The rabbit β-globin gene has been used for most studies of promoter function; the sequence of the human β-globin gene promoter7,14 suggests that it is quite similar in functional organization.

Four β-globin genes, isolated from individuals with β-thalassemia, have been shown to have single nucleotide substitutions in the promoter region (Fig. 2A).24–28 Transcription of three of these genes was markedly down compared to a normal gene in a cellular expression assay.25,27,28 These observations established that single nucleotide substitutions in the promoter region can cause the thalassemia phenotype in human ery-
thyroid cells. The definition of the β-globin gene promoter, derived from in vitro analysis, has thus been validated by study of these thalassemia globin genes.

The δ-globin gene may be considered “thalassemic” in that it encodes for a normal globin that is synthesized at a very low level. Functional studies have shown that the “defect” in the δ-gene can be localized to the promoter region. Note that the “CAT” box is imperfect and the distance between it and the TATA box is shorter than for functional genes (Fig. 2A). Furthermore, the δ-gene lacks the conserved sequences between 86 and 108 nucleotides upstream from the CAP site. Undoubtedly, this defective promoter structure largely accounts for the gene’s hypofunction, although recent studies have suggested that δ-mRNA is also somewhat less stable than β-mRNA in reticulocytes.

Remote DNA Sequences That Influence Gene Expression

Certain thalassemia mutations are characterized by deletion of large segments of the β-gene cluster (Fig. 3). Among the more interesting of these is one found in a family with γβ-thalassemia; deficient synthesis of each of these globins has been documented in affected individuals. This deletion has resulted in loss of more than 60 kilobases (kb) of DNA, including the γ and δ-genes, but has left the β-gene intact, along with 3 kb of upstream flanking sequences. This β-gene has been cloned. Its DNA sequence is normal and it functions normally in a cellular expression assay in vitro despite its lack of function in vivo. These observations provide evidence that sequences remote from the promoter can markedly influence the transcription of the globin gene in erythroid cells.

What are the possible mechanisms by which remote DNA sequences could influence promoter function? In cellular expression assays, the β-globin gene promoter requires an “enhancer” for function. The enhancer consists of a short segment of DNA that acts, independent of orientation or distance from the gene, to ensure accurate and efficient initiation of transcription. Such enhancers were first found in viral DNA, but recently, an enhancer element has been found in the introns of immunoglobulin genes. This enhancer is thought to activate the variable gene region promoter after the rearrangement that creates a functional immunoglobulin gene. Perhaps loss of function of the intact β-globin gene in the γδβ-deletion chromosome could reflect the loss of an essential enhancer of the β-gene promoter.

Alternatively, the large deletion could result in loss of DNA sequences that are important for establishing an active chromosomal conformation. Most DNA is associated with histone octamers (nucleosomes) and is tightly compacted into chromatin. Actively transcribed genes are in an “open” conformation that leaves them accessible for interaction with RNA polymerases and other proteins that may be involved in transcriptional regulation. Certain segments, referred to as hypersensitive sites because of their exquisite sensitivity to digestion by nucleases, are thought to consist of naked DNA that is unassociated with histones. The importance of such hypersensitive sites for promoter function has been shown for the Drosophila glue protein gene. Deletion of upstream
DNA that is normally found in hypersensitive sites has resulted in loss of function of the associated, structurally intact promoter. Similar remote sequences might also be involved in globin gene expression and, if missing, lead to hypofunction of a structurally normal gene. Indeed, the chromatin structure of the β-globin gene on the chromosome from which the ε, γ, and δ-genes have been deleted (Fig. 3C, γδβ-thal) lacks hypersensitive sites and is in a closed conformation in erythroid cells, where the β-gene on the normal chromosome has the characteristic structure of an expressed gene.

Also summarized in Fig. 3 are several other deletion mutations that influence expression of the γ-globin genes. These mutations have been carefully studied with the hope that they would provide insights into the mechanism(s) that regulate the developmental switch from fetal (HbF = α2γ2) to adult hemoglobin production (HbA = α2β2) during the perinatal period. The γδ-thalassemia mutations lead to a modest increase in HbF in adult life that is heterogeneously distributed among red cells (heterocellular).1 In contrast, the hereditary persistence of fetal hemoglobin (HPFH) mutations lead to a more striking increase in HbF, with a more even distribution among red cells (pancellular). Huisman and Schroeder first formulated the hypothesis that sequences between the γ and δ-genes might influence the developmental switch from HbF to HbA synthesis.59 Furthermore, they suggested that deletion of these sequences with HPFH mutations, but not in the δβ-thalassemia mutations, could account for their differing phenotypes. Discovery of members of the Alu class of moderately repetitive DNA sequences in this region focused attention on these as potential regulatory elements.60,61 Indeed, with exception of the Turkish thalassemia deletion that also removed the γ-gene,62 the presence of at least one of the two Alu sequences upstream from the δ-gene is associated with the δβ-thalassemia phenotype, whereas deletion of all or a portion of both Alu sequences is associated with the HPFH phenotype.62 However, the two classes of mutations differ with respect to size (Fig. 3), and hence, with respect to the nature of the DNA sequences brought into the γ-gene region. These factors might also be relevant to the different phenotypes of the two classes of mutations.

There are several HPFH mutations that lead to increased production of HbF in adults without altering the general structure of the β-like gene cluster (nondeletion HPFH).1 Certain of these mutations are not tightly linked to the β-like gene cluster,63,64 but others, particularly those that increase 6γ or 5γ synthesis selectively, may reflect sequence alteration in functionally important parts of these genes. Molecular studies in progress in several laboratories should ultimately lead to characterization of these interesting forms of HPFH.

Mutations That Affect RNA Processing

The study of RNA metabolism in erythroid cells suggested that many patients with β-thalassemia would have defects in RNA processing.65-68 This prediction has been proven correct by the sequence analysis and in vitro functional characterization of several cloned thalassemic globin genes.25,27,69-87 Indeed, the analysis of these globin genes has been remarkably fruitful in providing detailed insights about the RNA sequences required for specific, selective, and efficient RNA splicing.

The Consensus Splice Junction Sequences

Comparison of the sequences found at splice junctions provided the first clues as to the requirement for RNA processing, much in the way that study of promoter sequences identified conserved blocks that have subsequently been proven to be essential for promoter function. First to be identified were the dinucleotides, GT at the 5′ end and AG at the 3′ end of introns.88 A recent comparison of more than a hundred splice junctions has provided the expanded consensus splice junction sequence shown in Fig. 2B.15 The splice junction consensus sequences appear to be involved in the formation of a base paired structure between U1 RNA and the precursor RNA.89,90 U1 RNA is a small, stable, abundant, nuclear RNA species that may function as a cofactor in this splicing reaction. Adjacent portions of the 5′ end of U1 RNA are complementary to the 5′ and 3′ splice junction sequences; by forming a base paired structure with U1, the 5′ and 3′ ends of an intron are juxtaposed in a way that could readily facilitate accurate splicing.

Partial matches to the consensus sequences occur frequently at positions other than at the exon–intron boundaries. For example, GT occurs 44 times in the β-globin gene within a DNA segment that matches the 5′ consensus sequence at 5 of 9 positions, and AG occurs 49 times within a sequence that is a 9 of 16 match. These cryptic splice sites are rarely (if ever) used in normal RNA processing, but may be activated by mutations that make them a better match with the consensus sequence or by an effect of remote mutation. Study of various thalassemia mutant genes has yielded insight into the functional significance of individual nucleotides within the consensus sequences and has provided clues as to how splicing at one site may influence splicing at cryptic splice sites at other positions (Fig. 4).
THALASSEMIA

Mutations That Alter Splice Sites

Several mutations completely destroy the function of splice sites (Fig. 4A).\textsuperscript{24,25,27,52,82,87,91} Usually, these mutations lead to activation of otherwise cryptic splice sites in coding or intervening sequence RNA. One intriguing aspect of these mutations is the suggestion that there is some “pressure” to complete a splice within the globin gene transcript rather than leaving the RNA molecule unspliced.

Two $\beta^+$-thalassemia genes have been found to have a single nucleotide substitution in the consensus sequence region at the exon 1–intron I boundary, as illustrated in Fig. 4A.\textsuperscript{24,25} These mutant splice sites continue to function, as shown by the presence of normally spliced mRNA in test cells in vitro, but cause several abnormalities in mRNA processing. There is a quantitative decrease in the amount of correctly spliced mRNA, an increase in the amount of unspliced globin RNA, and activation of a cryptic splice site in coding sequence RNA. Thus, these consensus nucleotides, although not absolutely essential for splice site function, have a marked quantitative effect on RNA processing efficiency.

Mutations That Create an Alternate Site

Shown in Fig. 4B are several mutations that create alternate splice site. The first to be described involved replacement of G with A at a position 19 nucleotides upstream from the normal intron I–exon 2 boundary.\textsuperscript{70,71} This region of the normal transcript matches the 3' consensus splice sequence very well, except for lack of the essential dinucleotide, AG. The $\beta$-thalassemia mutation creates an AG, allowing the splice site to function. Indeed, 90% of the mRNA molecules transcribed from this $\beta$-thalassemia gene in vitro are incorrectly spliced at the alternate site.\textsuperscript{76,79} Such incorrectly spliced RNA molecules cannot be translated into $\beta$-globin.

Centered on the GT dinucleotide within codon 25 is a block that matches the 5' splice site consensus sequence in 6 of 9 positions. Nonetheless, this site is rarely if ever involved in splicing reactions during the processing of normal $\beta$-globin gene transcripts. However, this is one of the cryptic sites that is used when the normal exon 1–intron I boundary sequence is altered by mutation.\textsuperscript{25} Three separate mutations within this cryptic splice site also result in its activation.\textsuperscript{81,92,93} Substitution of A for T in codon 24 is silent at the level of protein sequence but leads to incorrect processing of about 80% of $\beta$-globin RNA molecules and thereby causes a moderately severe $\beta^+$-thalassemia phenotype.\textsuperscript{81} The other two mutations alter $\beta$-globin protein sequence and also lead to the use of this cryptic splice site, although less frequently than with the gene containing the codon 24 substitution.\textsuperscript{92} Hence, the $\beta^E$-gene and the Hb Knossos gene are associated with a mild $\beta$-thalassemia phenotype.

Mutations That Create an Alternate Site and Also Activate a Cryptic Site

Three mutations within intron II create alternate 5' splice sites.\textsuperscript{24,25,80,85,89a} These are located near the 3' end of the intron, as shown in Fig. 4C. One of the more interesting aspects of these mutations is the concurrent activation of a cryptic 3' splice site further upstream.
within the intron. This sequence matches the 3’ consensus splice sequence in 15 of 16 positions, but apparently is rarely, if ever, used during normal processing. The spliced product formed by use of this 3’ site within intron II and the normal 5’ site at the exon 2–intron II boundary cannot be spliced further, since there is not a 5’ splice sequence immediately downstream from the internal 3’ site in the transcript. At least one of these mutations appears to create a β0-gene, in that none of the β-globin RNA molecules found in test cells are processed to link exon 2 and exon 3 correctly.93

Splicing Mutations—Implications for RNA Processing Mechanism

A careful study of the various thalassemia mutations may contribute to formulation of a general mechanism of RNA processing. Several lessons have already been gained from the study of these mutant genes. For example, the dinucleotides, GT and AG, have been shown to be obligatory for function of splice sites, and the consensus nucleotides have a quantitative effect on splice site function. We have also learned that the splicing pattern of one intron may influence the removal of the other set of intervening sequences. For example, the alternate splice site mutation in intron I (Fig. 4B), in addition to affecting the splicing pattern, also retards removal of intervening sequences transcribed from both introns I and II.76,79 Similarly, the G to A substitution at the exon 2–intron II boundary retards processing of intron I RNA.82 This mutation also causes aberrant splicing, involving exon 1–intron I and intron II–exon 3 boundary sequences to link exons 1 and 3.

Finally, the mutations thus far described affect splice sites directly. RNA molecules form a complex secondary structure; such a secondary structure could influence splicing by bringing two splice sites in proximity or by “burying” a potential splice site within base paired portions of the RNA molecule.84 Indeed, study of mutant viral genes supports the influence of the secondary structure on the processing pattern. Perhaps activation of the cryptic 3’ site in intron II (Fig. 4C) by the alternate splice site mutations could reflect an alteration in the secondary structure of the RNA involving the cryptic site.

A Polyadenylation Signal Mutation

RNA transcription apparently continues beyond the position on DNA that corresponds to the 3’ end of mature mRNA molecules. For example, the mouse globin gene primary transcript has recently been shown to terminate roughly 750 nucleotides (nt) downstream from mRNA coding sequence.15 The sequence, AAUAAA, in the RNA transcript is thought to act as a recognition site for cutting the primary transcript some 20 nt downstream from the signal, thereby allowing addition of the poly A tract. In vitro mutagenesis of the corresponding DNA segment, AATAAA, does indeed affect polyadenylation.95 Recently, an α-globin gene, cloned from the DNA of an individual with HbH disease, has been shown to have a single nucleotide substitution in the polyadenylation signal (Fig. 2B).96 This leads to a quantitative reduction in α-mRNA and α-globin production in this patient who is homozygous for this mutation in the αs-gene and for a frameshift mutation in the αt-gene that abolishes gene function completely. In the transient expression assay in HeLa cells, the polyadenylation signal mutation leads to a marked reduction in correctly processed and polyadenylated RNA and gives rise to longer transcripts, extended at the 3’ end.

Mutations That Affect RNA Translation

There are several ways by which specific nucleotide substitutions could influence the efficiency with which RNA is translated into protein. However, the only mutations described to date are those that influence termination of protein synthesis. Specific codons (3 nucleotides) are used to signal the stop position of translation. Nucleotide substitutions that alter a codon that specifies an amino acid to a terminator codon cause premature termination of translation, whereas substitutions in the normal terminator allows synthesis of elongated globin chains.

Premature Termination Mutants

Several such mutations have been described.69,73,75,77,83 Each involves either a single nucleotide substitution in a normal codon to create a termination codon or a deletion that alters the reading frame of the mRNA to put a termination codon into phase. Examples of both are shown in Fig. 5. Note that several of the mutations that alter the splicing pattern of globin RNA also result in formation of mRNA molecules that have in-phase termination codons that lead to premature termination.70,77,79,92 Mutations that cause premature translation termination are sufficient to cause the thalassemia phenotype. These mutations are also associated with a variable, but usually a marked, quantitative decrease in the amount of globin mRNA.69,83,97,98 The traditional explanation for this quantitative effect of premature termination mutations is that the untranslated portion of the mRNA becomes accessible to nuclease degradation in the cytoplasm. However, recent studies suggest that premature termination mutations may affect intranuclear...
RNA stability or nuclear to cytoplasmic transport and therefore imply a previously unrecognized link between mRNA translation and nuclear RNA metabolism.99,100

**Terminator Codon Mutants**

The normal termination codon for both α and β-globin is UAA; α-globin mRNA contains 109 nucleotides in the 3' untranslated region, and β-mRNA contains 132 nucleotides. Replacement of a single nucleotide in the normal terminator to give a codon that specifies an amino acid rather than termination leads to synthesis of elongated globin chains. As indicated in Table 1B, several different substitutions in the α-mRNA terminator codon have been discovered; the resulting mutant hemoglobinss are structurally different only in the amino acid at the terminator codon position.103 The first to be discovered was Hb Constant Spring (CS). HbCS α-mRNA is read to the next in-phase terminator to give a “globin” with 172 rather than 141 amino acids. Shorter globin products of 169 and 154 amino acids apparently reflect limited proteolysis of the elongated α-globin. HbCS α-globin is present in red cells of affected individuals in very low quantities. Recently, the ability to measure α, and αγ-mRNA independently102,104 has revealed that the CS mutation occurs in the αγ-gene and that its mRNA is quantitatively deficient.103 This deficiency is thought to reflect instability of the HbCS α-mRNA,101 but this mechanism has not been directly demonstrated.

Not all elongated globins are associated with the thalassemia phenotype. For example, Hb Cranston contains a β-chain with 157 rather than 146 amino acids.105 Insertion of two nucleotides into codon 145 results in a shift in the mRNA translation reading frame; the normal terminator is out of phase, but a terminator codon in the new reading frame stops protein synthesis after addition of the 157th amino acid. Heterozygotes for the βCranston-gene have normal red cells and a normal HbA2. In contrast, heterozygotes for the βTak-gene (which also encodes a globin of 157 amino acids because of insertion of 2 nucleotides into codon 147) have hypochromic, microcytic red cells and normal or only slightly elevated HbA2 levels.106 Because of the different sites of insertion, βTak and βCranston differ in the amino acids found at positions 145 and 146; βCranston has the normal amino acids Tyr and His, respectively, while βTak has Ser and Ile. The molecular basis for the differing phenotypes of these mutants has not been elucidated.

**Alpha Thalassemias**

Alpha thalassemia is most often due to gene deletion, although with some mutations, the α-genes are
Alpha Gene Deletions

The advent of molecular cloning, DNA sequencing, and restriction endonuclease mapping techniques has led to an increased understanding of the molecular basis for the DNA rearrangements that cause α-thalassemia. DNA sequence analysis of the two α-globin genes and study of their flanking sequences has revealed several tandemly duplicated highly homologous segments within the α-gene cluster (Fig. 6).4,5

The presence of these highly homologous regions predisposes the α-globin gene cluster to recombinational events. Misalignment between two strands of DNA from two separate chromosomes during meiosis can effectively lead to deletion of one of the two α-globin genes. One type of deletion involves the homologous regions 3' to the αα-gene on one chromosome and 3' to the αβ-gene on the other and leads to deletion of 4.2 kb of DNA, whereas a second type involves the homologous segments containing the αβ and αα-genes and leads to deletion of 3.7 kb of DNA.109,110 These are referred to as the leftward and rightward deletions, respectively (Fig. 6). The outcome of these unequal crossovers is the generation of new chromosomes containing either a single α-gene or triplicated α-genes.109,123 All four possible chromosomes have been identified in various populations where α-thalassemia occurs in high frequency. Further support for this mechanism of unequal crossover is a similar occurrence observed in vitro during the propagation of cloned human DNA fragments containing the α-genes in E. coli. The deletions occurring in E. coli are indistinguishable from the two types of deletions described above. Note that the actual crossover sites could occur anywhere within the homologous segments and that there may be more than one mutational event that has led to single or triplicated α-gene chromosomes in each case.

The leftward type of deletion is prevalent in Oriental individuals122 and also occurs in low frequency in Saudi Arabia,123 but is never encountered in Mediterraneans or American blacks.117 The rightward deletion has a worldwide distribution among all racial groups (Orientals, Mediterraneans, and blacks).109-119 The single α-gene chromosomes are much more common than the triple α-gene chromosomes.41,120-123 Greek Cypriots exhibit the highest frequency of α-thalassemia among white people and also have a relative high frequency of triple α-gene chromosomes.124

Alpha globin production is roughly proportional to the number of functional α-globin genes,1 although slightly more α2 than α1-mRNA is produced in normal erythroid cells.102-104 Deletion of a single α-gene leads to an insignificant decrease in α-globin production in an individual having two functional α-globin genes on the other chromosome. Phenotypically, such individuals have been classified as having α-thalassemia-2 (silent carrier), and the single α-gene chromosome may be considered to be an αα-thalassemia mutation. Individuals with two chromosomes containing only a single α-globin gene have a more significant decrease in α-globin production; such patients have α-thalassemia-1 (mild hypochromic microcytic anemia) according to the standard clinical classification.1

---

Fig. 6. A summary of deletions in the α-globin gene cluster. The line diagram indicates the relative position of the genes and pseudogenes within the cluster. The two regions labeled “variable” are places where DNA length polymorphisms have been found; several different alleles, characterized by DNA fragments of different lengths in this region, have been defined.109-110 The positions of several tandemly duplicated DNA segments (X, Y, and Z) are indicated.103 The extent of the deleted segments for each of the mutations is shown. "Haplotype" in this context refers to the number of α-genes or fragments of a gene found on each chromosome, whereas the phenotype is that found in heterozygous individuals; α-thal-2 is a silent carrier state and α-thal-1 is thalassemia trait (mild hypochromic, microcytic anemia). The interested reader should consult an excellent recent review on the α-thalassemias for further details and specific references to each of these mutations.101
Several other deletions involve both α-globin genes of a single chromosome, as shown in Fig. 6.113,114,116-128 These are αα-thalassemia mutations. The extent of these deletions has been defined by restriction endonuclease mapping, but the mechanism of their occurrence is not yet obvious. Inheritance of a chromosome lacking both α-genes leads to the α-thalassemia-1 phenotype if the other chromosome is normal. If the other chromosome 16 has only a single α-gene, HbH disease (a moderately severe hemolytic anemia) occurs.

A novel and potentially important deletion in the α-like globin gene cluster was characterized in two patients of North European origin with an unusual combination of acquired HbH disease and mental retardation.129 This phenotype apparently occurred from the genetic combination of a standard “rightward” deletion chromosome inherited from the mother and a de novo deletion of the whole α-globin gene cluster on the other chromosome (Fig. 6). The father was clinically normal. This deletion must therefore represent an acquired mutation that occurred either in the germ cells of the father or at a very early stage of fetal development. A third patient also apparently inherited a de novo mutation from his father that inactivated both α-genes; this mutation was apparently of the nondeletion type.129 The combination of these unusual acquired mutations and mental retardation might be coincidental or could imply the presence of a locus within the α-gene cluster or within nearby DNA sequences that influences mental development.

Nondeletion α-Thalassemia Mutations

These comprise a distinct and apparently heterogeneous group of mutations that lead to absent or greatly reduced expression of α-genes that are structurally intact, as determined by restriction endonuclease mapping.112,113,130-133 The inheritance of such mutations, in combination with either α− or αα-type deletion mutations, leads to a highly variable clinical pattern.101 Undoubtedly, the nondeletion types of α-thalassemia represent the same types of mutations described in detail above in the genes of patients with β-thalassemia. Indeed, a 5-nucleotide deletion at a splice junction leads to abnormal splicing (Fig. 4A),98 and a second mutation (described below) leads to synthesis of a structurally abnormal globin (125 Leu → Pro) that is highly unstable.134 The interesting mutation, representing substitution of a single nucleotide in the polyadenylation signal (Fig. 2B), found in a nondeletion type α-thalassemia gene96 has already been described.

The rather rare and intriguing syndrome of acquired HbH disease is associated with a variety of myeloproliferative disorders and has been described in 34 patients.135 The α-globin genes are structurally intact in the erythroid cells of such patients, and furthermore, they were found to be hypomethylated, a characteristic of highly expressed genes.136 A marked reduction in transcription of all four α-globin genes has been demonstrated, but no reactivation of β-gene expression was observed.135,136 The ratio of α1 and α2-mRNAs was normal, implying that the mutation causing this disorder has suppressed all four α-genes in a balanced manner.136 This result is most consistent with a mutation affecting an intranuclear factor required for α-gene transcription.

ξ-Gene Rearrangements

Because of the high sequence homology between the ζ and the ξ-genes, it is not surprising that unequal crossovers have led to generation of chromosomes having single or triplicated ξ-genes.108,137 At least two different sized deletions have been defined, reflecting different recombinational events. Further studies are needed to define the expression levels of these rearranged genes and whether these rearrangements have any significant physiologic effects during embryonic or adult life.

USE OF RESTRICTION ENDONUCLEASE
POLYMORPHISMS TO STUDY BETA THALASSEMIA

The study of genomic variation at the level of DNA in man has led to immediate applications to population genetics, prenatal diagnosis, and characterization of various diseases. With molecular probes and a vast array of restriction endonucleases, a new body of information has accumulated concerning DNA sequence polymorphisms in the globin gene clusters, and their linkage to various specific thalassemia mutations and hemoglobinopathies has been established.

Polymorphisms are defined as genetic differences among individuals.138 The two possible alleles for a restriction endonuclease cleavage site are either (+), namely the enzyme cuts at that site, or (−), the enzyme does not cut. A particular site is considered to be polymorphic in a given population if the least frequent allele is present in more than 1% of individuals. The first restriction endonuclease site polymorphism found useful for study of hemoglobinopathies was that involving an Hpa I site, 4 kb downstream from the β-globin gene.139 In the first study, 87% of chromosomes having a sickle (βS) globin gene also lack this Hpa I cutting site (−), whereas the (−) allele is very infrequently found in DNA of individuals who do not have a βS-globin gene. This linkage of the Hpa I (−) allele to the βS mutation has been useful for prenatal diagnosis139,140 and for population genetic studies.141,142 Discovery of the Hpa I polymorphism and its use in sickle cell disease prompted a search for
other polymorphic sites within the \( \beta \)-globin gene cluster.\textsuperscript{144,151} However, the linkage of a specific allele at a single restriction endonuclease site with a particular globin gene mutation [such as the tight linkage of the Hpa I (\( \rightarrow \)) with the \( \beta^{s} \)-gene] is quite rare.

By using several polymorphic restriction sites to define the haplotype of an individual chromosome, Orkin and Kazazian and their colleagues have found an association of specific thalassemia mutations with particular haplotypes (Fig. 7).\textsuperscript{24} This was initially accomplished by cloning and sequencing globin genes from DNA of thalassemic individuals with each of the several haplotypes. Coupling of specific haplotypes to particular mutations has subsequently permitted restriction endonuclease analysis to be used for identification of specific mutations in individual patients.

The data shown in Fig. 7 apply only to the Mediterranean population. In other ethnic groups, e.g., American blacks and Asians, the same haplotypes are found to be associated with completely different mutations (Orkin and Kazazian, personal communication). This is consistent with genetic isolation of the various groups in whom thalassemia is common during the period when malaria operated as a major selective factor for the thalassemia trait phenotype. Study of new patient populations by haplotype determination provides an initial estimate of the number of different mutations in that population and allows examples of each to be selected for molecular characterization.

Coupling of a haplotype with a specific mutation is not absolute, even within a single ethnic group. In their initial study, Orkin et al. found the same mutation (\( \beta^{0.39} \)-premature termination codon) associated with two different haplotypes.\textsuperscript{24} Kan and colleagues have recently found that the \( \beta^{0.39} \)-mutation in the Sardinian population is associated with five different haplotypes.\textsuperscript{159} Another exception was the finding of more than one mutation associated with a single haplotype. These findings may imply that the same mutation has arisen more than once,\textsuperscript{160} or more likely, that DNA sequence rearrangements in the \( \beta \)-cluster occur with a far higher frequency than was heretofore suspected. From a practical perspective, haplotype analysis, although useful in identifying thalassemia mutations, is not absolutely accurate. Overall, the linkage of specific mutations to a particular haplotype holds about 80%-90% of the time.

A mutation that causes thalassemia or a hemoglobinopathy also may result in loss of a restriction endonuclease cutting site. For example, the \( \beta^{6} \)-mutation in the sixth codon of the \( \beta \)-globin gene results in loss of cutting sites for Mnl I, Dde I, and Mst II. Use of Mst II to recognize the \( \beta^{6} \)-mutation in amniotic fluid cell DNA has provided an easily applied and practical method for prenatal diagnosis.\textsuperscript{161-163} Among the thalassemia mutations that can be recognized by restriction endonuclease analysis are those that occur at \( -87 \) in the \( \beta \)-globin gene promoter (Fig. 2A—mutation I, Avr II),\textsuperscript{24} at the exon 1—intron I splice junction of the \( \alpha \)-globin gene (Fig. 4A—mutation 4, Hph I),\textsuperscript{72} at the exon 2—intron II splice junction of the \( \beta \)-globin-gene (Fig. 4A—Hph I),\textsuperscript{91} and within intron II of the \( \beta \)-globin gene (Fig. 4C—mutation 6, Rsa I).\textsuperscript{80} These mutations are relatively infrequent in individuals with \( \beta \)-thalassemia, so that direct restriction endonuclease analysis is not of great practical value.

**CLINICAL AND GENETIC HETEROGENEITY OF THE THALASSEMIAS**

The clinical heterogeneity of the thalassemias is based in part on the diversity of specific mutations and their quantitatively variable effect on mRNA and globin production. For example, individuals with two \( \beta^{0} \) mutations usually will have thalassemia major (transfusion-dependent). Similarly, individuals homozygous for the alternate splice site mutation in intron I (no. 5 in Fig. 4B and Fig. 7) also have severe disease because precursor mRNA molecules transcribed from the gene with this mutation are spliced incorrectly 90% of the time. Hence, patients homozygous for this defect have a marked quantitative deficiency of \( \beta \)-globin production.\textsuperscript{152} In contrast, an individual homozygous for a mutation in a consensus splice junction sequence
(mutation 2, Fig. 4A) that decreases, but does not abolish splicing, has milder disease—the thalassemia intermedia phenotype.\textsuperscript{132,132a}

In certain populations, such as the American Greek and Italian emigrants studied to obtain the data shown in Fig. 7, several mutations are present, some of which are found quite frequently. Hence, most patients with severe thalassemia in this group, although functionally homozygous in that both \( \beta \)-globin genes are defective, are, in fact, doubly heterozygous for two different mutations. In such populations, homozygous thalassemia will be extremely heterogeneous, and further studies correlating molecular genotype with clinical phenotype will be most informative. In other populations, only one mutation may be found and the thalassemia phenotype in homozygous individuals will be quite homogeneous. Both haplotype analysis, using restriction endonuclease polymorphisms,\textsuperscript{14} and direct DNA analysis, with oligonucleotide probes specific for a particular mutation,\textsuperscript{15b} will provide genotype information that can be correlated with clinical analysis.

Other modifiers of clinical phenotype in patients with homozygous thalassemia include coincident inheritance of an \( \alpha \)-thalassemia mutation\textsuperscript{153-156} or a mutation that increases HbF synthesis.\textsuperscript{157,158} Alpha thalassemia reduces disease severity, as the pathophysiology of severe \( \beta \)-thalassemia is based on the relative excess of \( \alpha \)-globin production. Any modifier that reduces this relative excess will be beneficial. Similarly, enhanced \( \gamma \)-globin synthesis reduces the imbalance in globin production and has a beneficial effect. Conversely, coinheritance of a chromosome with three \( \alpha \)-genes increases the severity of \( \beta \)-thalassemia.\textsuperscript{152c}

The \( \alpha \)-thalassemia syndromes are also extremely clinically heterogeneous. Detailed analysis of genotype has provided many insights into the molecular basis for this heterogeneity. The reader is directed to an excellent review on the \( \alpha \)-thalassemias by Higgs and Weatherall\textsuperscript{101} for a comprehensive consideration of the \( \alpha \)-thalassemia syndromes.

### THALASSEMIC HEMOGLOBINOPATHIES

Hemoglobinopathies are defined as abnormal hemoglobins with one or more structural changes, usually a single amino acid substitution in a globin chain that is synthesized at a normal or nearly normal rate. In contrast, thalassemia mutations are traditionally defined as those that leave the globin structure unchanged, but lead to a quantitative reduction in synthesis of the globin. The simple categorization of globin gene mutations has been violated by several that lead both to alteration of globin structure and to decreased synthesis. For example, as described above, the HbE [\( \beta(26) \) Glu→Lys]\textsuperscript{92} and Hb Knossos [\( \beta(27) \) Ala→Ser]\textsuperscript{93} mutations activate an alternate splice site, leading to a quantitative deficiency of \( \beta \)-globin mRNA (Fig. 4B). HbCS is another example where both structure and synthesis of \( \alpha \)-globin is altered by a single mutation.\textsuperscript{101,103} Similarly, the crossover globin, Hb Lepore,\textsuperscript{1} arises from a recombinational deletion that created a 5'\( \beta \) 3'\( \beta \) hybrid gene (Fig. 3), the product of which is synthesized at a very low rate. Hence, the Lepore gene causes thalassemia. Several other examples may be cited, as shown in Table 1. These mutations have been reviewed in detail recently.\textsuperscript{164} Only those in which the mechanism of the thalassemic effect has been at least partially clarified will be discussed here.

The \( \beta \)-chain found in Hb Indianapolis has been shown to be highly unstable by biosynthetic studies.\textsuperscript{165} The \( \beta\textsuperscript{Ind} \)-chain apparently preferentially associates with the red cell membrane, where it is also catabolized, as shown by pulse-chase studies. Hb Indianapolis is unusual in that it causes the severe thalassemia phenotype in heterozygous individuals, perhaps due to membrane damage caused by the \( \beta\textsuperscript{Ind} \)-chain. Other unstable hemoglobin variants may be associated with mild hypochromia, but ineffective erythropoiesis and severe anemia (found in individuals with Hb Indianapolis) is not usually observed. Heterozygosity for a mild \( \beta \)-thalassemia gene on the chromosome opposite the \( \beta\textsuperscript{Ind} \)-gene, suggested by increased HbF levels (10% and 12%), could explain the clinical severity. However, the affected individuals are father and daughter and there is no other evidence of \( \beta \)-thalassemia in this family.

The Hb Quong Sze mutation [\( \alpha(125) \) Leu→Pro] was discovered upon sequencing of the two \( \alpha \)-genes from one chromosome of an individual with HbH disease.\textsuperscript{134} Both \( \alpha \)-genes were missing from the other chromosome. One of the two sequenced \( \alpha \)-genes was normal. Discovery of the sequence abnormality in the other gene prompted the hypothesis that \( \alpha\textsuperscript{Quong Sze} \)-chain would be unstable, as leucine to proline substitutions often have a profound effect on the stability of the hemoglobin tetramer. This hypothesis was supported by biosynthetic studies that demonstrated a normal synthesis but instability of the mutant \( \alpha \)-globin.\textsuperscript{134}

Direct measurements have shown that the \( \beta \)-mRNA for Hb Vicksburg [\( \beta(75) \) Leu→0] is reduced in concentration,\textsuperscript{166} whereas deficiency of the \( \beta \)-mRNA for Hb North Shore [\( \beta(134) \) Val→Glu] has been inferred by biosynthetic studies.\textsuperscript{167} The deletion causing \( \beta\textsuperscript{Vicksburg} \) and the substitution causing \( \beta\textsuperscript{North Shore} \) do not alter \( \beta \)-globin RNA sequence so as to create or activate an alternate splice site. The reason for the thalassemic effects of these mutations therefore remains obscure.
Perhaps two mutations are present in each β-gene: one responsible for the structural change and the other having a thalassemic effect. Alternatively, sequence alterations in β-globin RNA may alter transport or processing by mechanisms not yet appreciated.

THE TREATMENT OF THALASSEMIA

The treatment of thalassemia pertains only to that of severe (or homozygous) β-thalassemia. Those patients with thalassemia trait (heterozygous α- or β-thalassemia) require no treatment. Patients with HbH disease also usually do quite well with no specific treatment. Because of their severe anemia, which requires lifelong transfusional support, nonchelated patients with severe β-thalassemia die from iron overload between the ages of 15 and 25 yr.

Conventional Therapy

Transfusions

Red cell transfusions were originally given to thalasemic patients to sustain life and then to relieve the symptoms of anemia. "Hypertransfusion," a strategy designed to maintain a minimal hemoglobin of 10 g/dl and a mean of 12 g/dl, was justified despite increased blood requirements because of improved growth and development during the first decade. In addition, maintaining the hemoglobin at higher levels tended to reduce the hyperabsorption of dietary iron that occurs with the dyserythropoietic syndromes. Splenectomy when the red cell requirement exceeds 250 ml/kg/yr is also a part of a standard conventional therapy.

With hypertransfusion, however, the bone marrow mass and blood volume remain expanded. For this reason, a more aggressive program of transfusions ("supertransfusion") was proposed. Maintenance of a mean hemoglobin of 14–15 g/dl will almost completely suppress endogenous erythropoiesis, thereby shrinking the mass of the bone marrow and reducing the blood volume. Red cell transfusion requirements therefore might be no greater in supertransfused patients than in hypertransfused patients because of this reduction in blood volume. This prediction appeared to be supported in a large study of several patients, but as yet, there is insufficient clinical experience to judge the long-term value of the "supertransfusion" regimen.

The red cell requirements of thalassemic patients could be reduced if young red cells ("neocytes") with prolonged survival in vivo could be given. Collection of neocytes by continuous flow centrifugation, or by fractionation of conventionally collected blood units with a cell washer, is feasible. Prolonged survival of such young red cells can be demonstrated by measurement of the 51Cr half-life. Well documented studies are needed to quantify the decrease in red cell requirement realized by use of neocyte units. A significant reduction in transfused iron must be demonstrated to justify the added effort and expense required to obtain young red cells.

Chelation Therapy

Desferrioxamine (Desferal) is the only iron chelator that is currently in widespread clinical use. Its role in the treatment of transfusional iron overload has recently been reviewed. This drug is effective in inducing iron excretion and is relatively nontoxic; unfortunately, it must be delivered as a continuous 8–12 hr subcutaneous infusion to be effective. Iron balance can be maintained in most patients with infusions of 1,500–2,000 mg of the drug given 5 or 6 days a week. Supplemental intravenous infusions of larger amounts of the drug can be used to markedly augment iron removal, particularly in older patients with marked iron overload. A decrease in liver iron concentration, an improvement in liver function tests, and reduced ferritin levels have been demonstrated in older patients on intensive desferrioxamine treatment. Attainment of iron balance in younger patients may be demonstrated by comparing transfused iron to iron that is excreted. Iron balance is reflected by low serum ferritin levels.

Prevention of cardiac damage by desferrioxamine has not been demonstrated, however. Propper et al. have recently reported that initiation of chelation in thalassemic patients after the age of 10 does not appear to prevent cardiac disease. However, we have observed an adult with acquired pure red cell aplasia in whom regular chelation was started shortly after the transfusion requirement was established. This patient reached an iron burden equivalent of 300 U of packed red cells without significant cardiac iron deposition. At present, despite the expense and cumbersome nature of administration, the regular use of Desferal is the only hope of forestalling the consequences of iron overload in regularly transfused thalassemic patients.

Vitamin C is commonly employed to augment iron removal in response to desferrioxamine. The ability of this substance to mobilize iron from nontoxic stores, or to enhance lipid peroxidation by iron in myocardial cells, can apparently lead to cardiac deterioration in heavily iron overloaded individuals treated with vitamin C. Hence, a low dose should be used, the drug should not be given without a concurrent desferrioxamine infusion, and patients should be advised to avoid large quantities of citrus fruits or vitamin C supplements. Vitamin E deficiency, often present because of the excessive catabolism of this substance in iron overload.
overloaded individuals, may be corrected by oral administration.\(^{179}\) This reduces the amount of a peroxidation product (malonyldialdehyde) in red cell membranes, but prolongation of red cell survival has been demonstrated only when vitamin E is given parenterally to thalassemic patients.\(^{179}\)

**Development of Experimental Chelators**

The National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases and the National Heart, Lung, and Blood Institute have sponsored efforts to develop new chelating agents that are active orally. These efforts have led to identification of a drug, ethylene diamine-\(N,N^1\)-bis(2-hydroxyphenyl acetic acid)-dimethyl ester dihydrochloride (EDHPA), that has considerable ability to chelate iron in animals when administered orally (David Badman and W.F. Anderson, personal communication). Toxicity studies in animals have been initiated to determine whether these compounds are suitable for human use. Efforts to modify desferrioxamine chemically to make it active by the oral route have been unsuccessful. All other investigational chelating agents (e.g., rhodotorulic acid and 2,3-dihydrobenzoic acid) have been found to be excessively toxic or to be inefficacious in human trials.\(^{174}\)

**New Approaches to Treatment**

**Bone Marrow Transplantation**

Although transplantation from an HLA identical sibling involves an approximately 30% risk of death, successful transplantation of individual thalassemic patients can offer the chance for cure and perhaps a normal life expectancy.\(^{180,181}\) On the other hand, conventional therapy with regular blood transfusions can be expected to provide 12–15 yr of fairly good quality life before the complications of iron overload begin to develop. Death usually occurs between the age of 15 and 25 yr.\(^{174}\) The impact of chelation therapy on survival is not certain.\(^{175,181a}\) Nonetheless, there is the possibility that a newborn or very young child with thalassemia might benefit from new developments in genetic therapy (see below). Thus, the choice between bone marrow transplantation and conventional therapy represents a fundamental ethical dilemma, as the significant risk of death or chronic graft-versus-host disease must be weighed against the probability of cure in an individual patient.

To achieve a high probability of engraftment and to reduce the risk of graft-versus-host disease, the Seattle team has chosen to transplant very young thalassemic children who have received minimal blood transfusions prior to transplantation ("good-risk" patients).\(^{181,182}\) Five patients transplanted to date were conditioned with dimethyl busulfan and cyclophosphamide. Three patients achieved engraftment with cure of their thalassemia and have done well during follow-up of 11 mo to 2 yr, although two have died of transplant complications (E. Donnell Thomas, personal communication). Lucarelli and his colleagues in Pesaro have also achieved a high rate of engraftment in several good-risk patients using the same conditioning regimen, although their earlier efforts with a conditioning regimen including cyclophosphamide and total body irradiation (but no busulfan) were less successful. “High-risk” patients—older individuals with an extensive transfusion history—had a low frequency of engraftment (1 in 7) (Guido Lucarelli, personal communication), with a conditioning regimen that included cyclophosphamide and total body irradiation. From experience with bone marrow transplantation for conventional indications, it is highly likely that these patients, because of prior sensitization, are at a relatively high risk for graft rejection.\(^{180}\) About 25% of all transplanted thalassemia patients had significant graft-versus-host disease; and several patients have died of complications related to bone marrow transplantation.

Despite the small number of transplants thus far performed, the data suggest that some young, minimally transfused thalassemic patients may be cured of their disease by successful transplantation; others will die from complications of the transplant procedure. Only patients with HLA identical donors are currently candidates. Whether engraftment can be achieved in more heavily transfused patients remains to be determined. Transplantation for this disease should still be considered experimental and its application limited to highly experienced clinical investigators.

**Activation of the \(\gamma\)-Globin Genes**

The clinical severity of \(\beta\)-thalassemia is roughly proportional to the imbalance in \(\alpha\) and non-\(\alpha\)-globin synthesis.\(^1\) Those patients who also inherit a gene that increases \(\gamma\)-globin synthesis during postnatal life (HPFH)\(^{157,158}\) have a milder clinical course than the usual patient with \(\beta\)-thalassemia. Hence, reactivation of the developmentally repressed \(\gamma\)-globin genes has long been proposed as a therapeutic goal for treatment of patients with severe \(\beta\)-thalassemia.

Demonstration that cytosine residues near the \(\gamma\)-globin gene become methylated coincident with their developmental repression\(^{183,184}\) provided the rationale for a strategy designed to reactivate these genes.\(^{185,186}\) 5-Azacytidine is an analog of cytidine that has nitrogen rather than carbon in the fifth position of the pyrimidine ring; this is the position that is methylated
following DNA synthesis. Hence, 5-azacytidine, when incorporated into DNA during cell division, cannot undergo methylation and, more importantly, 5-azacytidine residues in the newly synthesized DNA strand apparently bind and inactivate the methyltransferase responsible for adding methyl groups to other sites. All DNA in cells exposed to 5-azacytidine, therefore, becomes hypomethylated after a few cell divisions. As 5-azacytidine had been shown to reactivate repressed genes in tissue culture cells, DeSimone, Heller, and colleagues administered the compound to baboons in an effort to reactivate the fetal globin genes. These investigators found a striking increase in γ-globin synthesis and a reciprocal decrease in β-globin synthesis in these animals, whose globin gene organization and structure is similar to that of man.

These observations have prompted limited trials of 5-azacytidine in patients with severe β-thalassemia and sickle cell disease. Four to seven-fold increases in γ-globin synthesis have consistently been observed. In the thalassemic patients, an increase in the reticulocyte count and hemoglobin level suggested that increased γ-globin synthesis led to more effective erythropoiesis and improved red cell survival. The effect of a 7-day course of 5-azacytidine disappears after 2–3 wk, so that regular or cyclic administration would be required to achieve a long-term increase in fetal globin synthesis.

5-Azacytidine administration leads to generalized hypomethylation of bone marrow cell DNA, but only the γ-genes are activated. The embryonic β-globin genes are also hypomethylated after treatment, but very little mRNA for this globin was found in the bone marrow cells of treated patients. Several other genes, including those for insulin, collagen, and the myc cellular proto-oncogene, also become hypomethylated without increased expression (T.J. Ley, unpublished observation). The tendency of 5-azacytidine to selectively reactivate the γ-globin gene (but not other genes on human chromosome 11) has also been demonstrated in vitro, using a somatic cell hybrid line derived from mouse erythroleukemia cells and human fibroblasts. The remarkable specificity of 5-azacytidine in activating the γ-genes, nonetheless, invites the suggestion that the mechanism by which it activates this gene is unrelated to its demethylating effect, but rather is due to a subclinical cytotoxic effect on erythroid progenitors. Bone marrow ablation with subsequent regeneration is regularly accompanied by a transient increase in HbF synthesis. The initial experimental studies in baboons suggested that drugs other than those that cause demethylation did not lead to γ-gene activation, but others have reported that cytosine arabinoside or hydroxyurea can lead to increased HbF production in primates. Thus, the mechanism(s) by which 5-azacytidine increases HbF production is not yet established.

5-Azacytidine is a carcinogenic compound, although its administration to man or experimental animals does not lead acutely to development of tumors. No tumors were observed after several months of administration of the drug to mice and rats, although many of the older animals did ultimately develop hematopoietic neoplasms. One study, involving a small number of Fischer rats, demonstrated an increased incidence of tumors when compared to control animals several months after the drug was stopped (Brian Carr, personal communication). More than 1,000 patients have received the drug as treatment for leukemia; no secondary tumors have yet been observed, although the average follow-up period is relatively short (information provided by the Investigational Drug Branch, National Cancer Institute).

The long-term carcinogenic risk has led some to argue that 5-azacytidine should not be given to thalassemic patients. Many questions about its mechanism and relative potency, compared to other drugs, in increasing HbF synthesis can be resolved by animal studies. Hence, human studies can be limited to those designed to determine whether the drug can eliminate or reduce the transfusion requirement of individual patients. The long-term risk of the drug, in such patients, should be weighed against the patient’s prognosis with conventional treatment.

Gene Therapy

Recombinant DNA methodology and the availability of cloned globin genes have provided the first realistic opportunity to consider approaches to functionally correct thalassemic defects by introducing new genetic information into bone marrow cells. Transfer of genes into tissue culture cells has become a routine procedure. Cells deficient in an enzyme, e.g., thymidine kinase, can be transformed to thymidine kinase expression by introduction of purified DNA fragments containing the thymidine kinase gene. Because the efficiency of transformation is low, a selective growth media must be used to inhibit growth of enzyme-deficient cells, while allowing growth of the transformed cells. Transformation of metabolically normal cells can also be achieved by use of genes for selectable markers that confer a genetic advantage on such cells in the presence of a particular drug. For example, the gene for a mutant, methotrexate-insensitive, dihydrofolate reductase has been cloned and will (if expressed) allow cells to grow in high concentrations of methotrexate.

Globin genes do not confer a growth advantage.
Hence, their introduction in cells requires cotransformation: a gene for globin and one for a selectable marker are introduced simultaneously. Using this strategy, the human β-globin gene has been introduced into mouse erythroleukemia cells, and several transformed clones have been shown to express the human β-globin gene.50 Exposure of these transformed erythroleukemia cells to the inducer, dimethylsulfoxide, leads to increased human β-globin gene expression as the cells mature into erythroblasts, in parallel with increased expression of the mouse globin genes. Although regulation of the human globin gene is apparently “normal,” not all clones express the human gene, and the level of its expression is considerably lower than that of the native mouse globin genes.

The general strategy for introducing globin genes into mouse hematopoietic stem cells was developed by Cline et al.192 Suspensions of hematopoietic cells were exposed to DNA in vitro, reinfused into lethally irradiated animals, and the transformed cells selected by treatment of the recipient animals with methotrexate. These experiments employed genomic DNA from cells that produced a methotrexate-insensitive dihydrofolate reductase. The availability of the cloned gene sequences for this enzyme196 should facilitate attempts at gene transfer into hematopoietic stem cells of experimental animals and allow cotransformation of such cells with globin genes physically linked to the selectable marker gene. Such linkage, accomplished by recombinant DNA techniques, increases the probability of stable integration—and perhaps of expression—of the newly introduced globin gene.

An alternate approach for functional correction of premature termination mutations is the introduction of a gene for a suppressor tRNA.198 Such tRNAs “read” termination codons and can be tailored to insert an appropriate amino acid into globin. It is not yet clear whether suppressor tRNA genes could be expressed at an adequate level or whether suppression of normal terminators would have deleterious effects on cells.

Many problems remain to be solved, in addition to the variable and low level of globin gene expression in transformed erythroid cells, before gene transfer can be proposed with a reasonable probability of success. Transformation frequency, still very low for tissue culture cells, may be increased substantially by use of retroviral vectors.199 Recombinant retroviral genomes, appropriately packaged into viral coat proteins, may be introduced into target cell populations with very high efficiency. Whether cotransformation of cells with both the gene for a selectable enzyme marker and that for globin can be accomplished with these retroviral vectors has not yet been ascertained.

The ethical issues of gene therapy have been extensively discussed.200–202 Certainly, one could not reasonably attempt to introduce genes into human bone marrow cells for therapeutic purposes until a procedure had been shown to be safe, reproducible, and effective in appropriate experimental animals. Despite the significant problems that remain to be solved, this approach does hold promise and may yet provide a definite method of treatment for the severe β-thalassemias.

ACKNOWLEDGMENT

We wish to thank Drs. E. Donnell Thomas, G. Lucarelli, and B. Carr for communication of unpublished results. We are grateful to Cathy Magruder and Linda Garza for help in preparation of this manuscript.

REFERENCES

9. Slightom JL, Blechl AE, Smithies O: Human fetal g2- and g2-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21:627, 1980
17. Proudfoot NJ, Brownlee GG: 3' Non-coding region sequences in eukaryotic messenger RNA. Nature 263:211, 1976


100. Markham AF, Giardina PJV, Li A, Kazazian HH: \(\beta\)-thalassemia globin mRNA deficient \(\beta\)-globin mRNA precursor to mRNA is defective in three patients with \(\beta\)-thalassemia. Proc Natl Acad Sci USA 77:4287, 1980


117. Higgs DR, Goodbourn SEY, Lamb J, Clegg JB, Weatherell...
in preleukemia is due to decreased expression of all four \(\alpha\)-globin genes. Proc Natl Acad Sci USA 80:6051, 1983


144. Jeffreys AJ: DNA sequence variants in the \(\alpha\)\textsubscript{y}, \(\alpha\textsubscript{y}\), \(\beta\)- and \(\beta\)-globin genes of man. Cell 18:1, 1979


183. van der Ploeg LHT, Flavell RA: DNA methylation in the human $\gamma$g-globin locus in erythroid and nonerythroid tissues. Cell 19:947, 1980
188. Charache S, Dover G, Smith K, Talbot C Jr, Moyer M, Boyer S: Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with non-random hypomethylation of DNA around the $\gamma$-b$\beta$ gene cluster. Proc Natl Acad Sci USA 80:4842, 1983
190. Ley TJ, Chiang YL, Haidaris D, Anagnou NP, Wilson V, Anderson WF: 5-Azacytidine selectively activates the human $\gamma$ globin gene expression in mouse erythroleukemia cells containing human chromosome 11. (submitted for publication)
193. Bioassay of 5-azacytidine for possible carcinogenicity, CAS no. 320-67-2; NCI-CG-TR-42; DHN/PUB/NIH-78-842. Bethesda, MD, National Cancer Institute, Carcinogenesis Program
Advances in thalassemia research

AW Nienhuis, NP Anagnou and TJ Ley