Chapter 1 Beginnings: the molecular pathology of hemoglobin

David Weatherall

Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK

Historical background

Linus Pauling first used the term "molecular disease" in 1949, after the discovery that the structure of sickle cell hemoglobin differed from that of normal hemoglobin. Indeed, it was this seminal observation that led to the concept of *molecular medicine*, the description of disease mechanisms at the level of cells and molecules. However, until the development of recombinant DNA technology in the mid-1970s, knowledge of events inside the cell nucleus, notably how genes function, could only be the subject of guesswork based on the structure and function of their protein products. However, as soon as it became possible to isolate human genes and to study their properties, the picture changed dramatically.

Progress over the last 30 years has been driven by technological advances in molecular biology. At first it was possible only to obtain indirect information about the structure and function of genes by DNA/DNA and DNA/RNA hybridization; that is, by probing the quantity or structure of RNA or DNA by annealing reactions with molecular probes. The next major advance was the ability to fractionate DNA into pieces of predictable size with bacterial restriction enzymes. This led to the invention of a technique that played a central role in the early development of human molecular genetics, called *Southern blotting* after the name of its developer, Edwin Southern. This method allowed the structure and organization of genes to be studied directly for the first time and led to the definition of a number of different forms of molecular pathology.

Molecular Hematology, 3rd edition. Edited by Drew Provan and John Gribben.

Once it was possible to fractionate DNA, it soon became feasible to insert the pieces into vectors able to divide within bacteria. The steady improvement in the properties of cloning vectors made it possible to generate libraries of human DNA growing in bacterial cultures. Ingenious approaches were developed to scan the libraries to detect genes of interest; once pinpointed, the appropriate bacterial colonies could be grown to generate larger quantities of DNA carrying a particular gene. Later it became possible to sequence these genes, persuade them to synthesize their products in microorganisms, cultured cells or even other species, and hence to define their key regulatory regions.

The early work in the field of human molecular genetics focused on diseases in which there was some knowledge of the genetic defect at the protein or biochemical level. However, once linkage maps of the human genome became available, following the identification of highly polymorphic regions of DNA, it was possible to search for any gene for a disease, even where the cause was completely unknown. This approach, first called *reverse genetics* and later rechristened *positional cloning*, led to the discovery of genes for many important diseases.

As methods for sequencing were improved and automated, thoughts turned to the next major goal in this field, which was to determine the complete sequence of the bases that constitute our genes and all that lies between them: the Human Genome Project. This remarkable endeavor was finally completed in 2006. The further understanding of the functions and regulation of our genes will require multidisciplinary research encompassing many different fields. The next stage in the Human Genome Project, called *genome annotation*, entails analyzing the raw DNA sequence in order to determine its biological significance. One of the main ventures in the era of functional genomics will be in what is termed *proteomics*, the large-scale analysis of the protein

products of genes. The ultimate goal will be to try to define the protein complement, or proteome, of cells and how the many different proteins interact with one another. To this end, large-scale facilities are being established for isolating and purifying the protein products of genes that have been expressed in bacteria. Their structure can then be studied by a variety of different techniques, notably X-ray crystallography and nuclear magnetic resonance spectroscopy. The crystallographic analysis of proteins is being greatly facilitated by the use of X-ray beams from a synchrotron radiation source.

In the last few years both the utility and extreme complexity of the fruits of the genome project have become apparent. The existence of thousands of single-nucleotide polymorphisms (SNPs) has made it possible to search for genes of biological or medical significance. The discovery of families of regulatory RNAs and proteins is starting to shed light on how the functions of the genome are controlled, and studies of acquired changes in its structure, *epigenetics*, promise to provide similar information. However, a full understanding of the interactions of these complex regulatory systems, presumably by major advances in systems biology, is still a long way in the future.

During this remarkable period of technical advance, considerable progress has been made toward an understanding of the pathology of disease at the molecular level. This has had a particular impact on hematology, leading to advances in the understanding of gene function and disease mechanisms in almost every aspect of the field.

The inherited disorders of hemoglobin – the thalassemias and structural hemoglobin variants, the commonest human monogenic diseases – were the first to be studied systematically at the molecular level and a great deal is known about their genotype–phenotype relationships. This field led the way to molecular hematology and, indeed, to the development of molecular medicine. Thus, even though the genetics of hemoglobin is complicated by the fact that different varieties are produced at particular stages of human development, the molecular pathology of the hemoglobinopathies provides an excellent model system for understanding any monogenic disease and the complex interactions between genotype and environment that underlie many multigenic disorders.

In this chapter I consider the structure, synthesis and genetic control of the human hemoglobins, describe the molecular pathology of the thalassemias, and discuss briefly how the complex interactions of their different genotypes produce a remarkably diverse family of clinical phenotypes; the structural hemoglobin variants are discussed in more detail in *Chapter 15*. Readers who wish to learn more about the methods of molecular genetics, particularly as applied to the study of hemoglobin disorders, are referred to the reviews cited at the end of this chapter.

The structure, genetic control and synthesis of normal hemoglobin

Structure and function

The varying oxygen requirements during embryonic, fetal and adult life are reflected in the synthesis of different structural hemoglobins at each stage of human development. However, they all have the same general tetrameric structure, consisting of two different pairs of globin chains, each attached to one heme molecule. Adult and fetal hemoglobins have α chains combined with β chains (Hb A, $\alpha_2\beta_2$), δ chains (Hb A_2, $\alpha_2\delta_2$) and γ chains (Hb F, $\alpha_2\gamma_2$). In embryos, α-like chains called ζ chains combine with γ chains to produce Hb Portland ($\zeta_2\gamma_2$), or with ε chains to make Hb Gower 1 ($\zeta_2\varepsilon_2$), while α and ε chains form Hb Gower 2 ($\alpha_2\varepsilon_2$). Fetal hemoglobin is heterogeneous; there are two varieties of γ chain that differ only in their amino acid composition at position 136, which may be occupied by either glycine or alanine; γ chains containing glycine at this position are called $^{G}\gamma$ chains, those with alanine $^{A}\gamma$ chains (Figure 1.1).

The synthesis of hemoglobin tetramers consisting of two unlike pairs of globin chains is absolutely essential for the effective function of hemoglobin as an oxygen carrier. The classical sigmoid shape of the oxygen dissociation curve, which reflects the allosteric properties of the hemoglobin molecule, ensures that, at high oxygen tensions in the lungs, oxygen is readily taken up and later released effectively at the lower tensions encountered in the tissues. The shape of the curve is quite different to that of myoglobin, a molecule that consists of a single globin chain with heme attached to it, which, like abnormal hemoglobins that consist of homotetramers of like chains, has a hyperbolic oxygen dissociation curve.

The transition from a hyperbolic to a sigmoid oxygen dissociation curve, which is absolutely critical for normal oxygen delivery, reflects cooperativity between the four heme molecules and their globin subunits. When one of them takes on oxygen, the affinity of the remaining three increases markedly; this happens because hemoglobin can exist in two configurations, deoxy(T) and oxy(R), where T and R represent the tight and relaxed states, respectively. The T configuration has a lower affinity than the R for ligands such as oxygen. At some point during the addition of oxygen to the hemes, the transition from the T to the R configuration occurs and the oxygen affinity of the partially liganded molecule increases dramatically. These allosteric changes result from interactions between the iron of the heme groups and various bonds within the hemoglobin tetramer, which lead to subtle spatial changes as oxygen is taken on or given up.

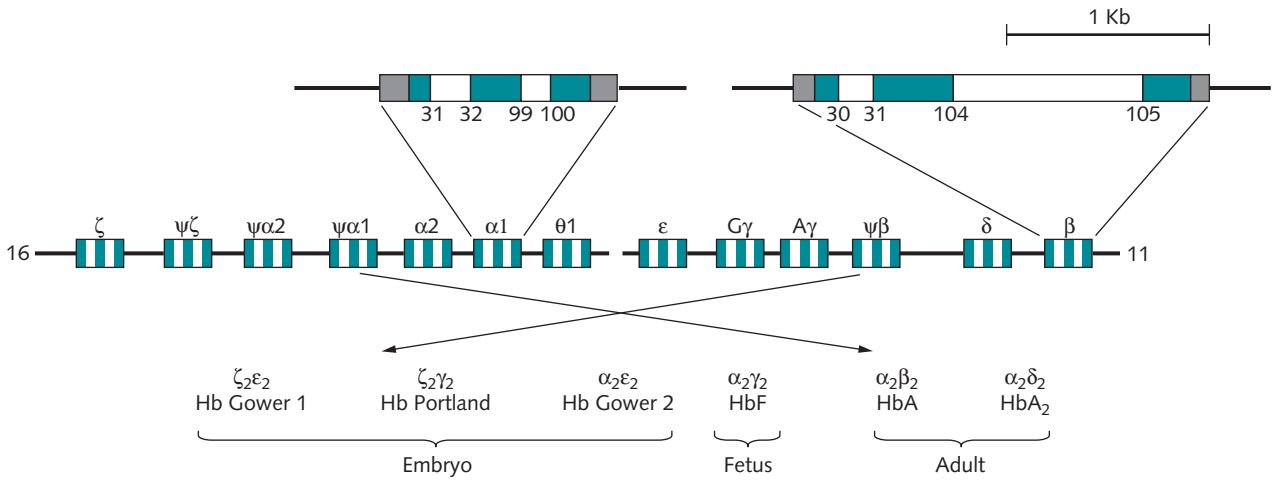

Fig. 1.1 The genetic control of human hemoglobin production in embryonic, fetal and adult life

The precise tetrameric structures of the different human hemoglobins, which reflect the primary amino acid sequences of their individual globin chains, are also vital for the various adaptive changes that are required to ensure adequate tissue oxygenation. The position of the oxygen dissociation curve can be modified in several ways. For example, oxygen affinity decreases with increasing CO_2 tension (the Bohr effect). This facilitates oxygen loading to the tissues, where a drop in pH due to CO_2 influx lowers oxygen affinity; the opposite effect occurs in the lungs. Oxygen affinity is also modified by the level of 2,3-diphosphoglycerate (2,3-DPG) in the red cell. Increasing concentrations shift the oxygen dissociation curve to the right (i.e., they reduce oxygen affinity), while diminishing concentrations have the opposite effect. 2,3-DPG fits into the gap between the two β chains when it widens during deoxygenation, and interacts with several specific binding sites in the central cavity of the molecule. In the deoxy configuration the gap between the two β chains narrows and the molecule cannot be accommodated. With increasing concentrations of 2,3-DPG, which are found in various hypoxic and anemic states, more hemoglobin molecules tend to be held in the deoxy configuration and the oxygen dissociation curve is therefore shifted to the right, with more effective release of oxygen.

Fetal red cells have greater oxygen affinity than adult red cells, although, interestingly, purified fetal hemoglobin has an oxygen dissociation curve similar to that of adult hemoglobin. These differences, which are adapted to the oxygen requirements of fetal life, reflect the relative inability of Hb F to interact with 2,3-DPG compared with Hb A. This is because the γ chains of Hb F lack specific binding sites for 2,3-DPG.

In short, oxygen transport can be modified by a variety of adaptive features in the red cell that include interactions between the different heme molecules, the effects of CO_2 and differential affinities for 2,3-DPG. These changes, together with more general mechanisms involving the cardiorespiratory system, provide the main basis for physiological adaptation to anemia.

Genetic control of hemoglobin

The α- and β-like globin chains are the products of two different gene families which are found on different chromosomes (Figure 1.1). The β-like globin genes form a linked cluster on chromosome 11, spread over approximately 60 kb (kilobase or 1000 nucleotide bases). The different genes that form this cluster are arranged in the order 5′–ε–^{G}γ–^{A}γ–ψβ–δ–β–3′. The α-like genes also form a linked cluster, in this case on chromosome 16, in the order 5′–ζ–ψζ–ψα1–α2–α1–3′. The ψβ, ψζ and ψα genes are pseudogenes; that is, they have strong sequence homology with the β, ζ and α genes but contain a number of differences that prevent them from directing the synthesis of any products. They may reflect remnants of genes that were functional at an earlier stage of human evolution.

The structure of the human globin genes is, in essence, similar to that of all mammalian genes. They consist of long strings of nucleotides that are divided into coding regions, or exons, and non-coding inserts called *intervening sequences* (IVS) or introns. The β-like globin genes contain two introns, one of 122–130 base pairs between codons 30 and 31 and one of 850–900 base pairs between codons 104 and 105 (the exon codons are numbered sequentially from the 5′ to the 3′ end of the gene, i.e., from left to right). Similar, though smaller, introns are found in the α and ζ globin genes. These introns and exons, together with short non-coding sequences at the 5′ and 3′ ends of the genes, represent

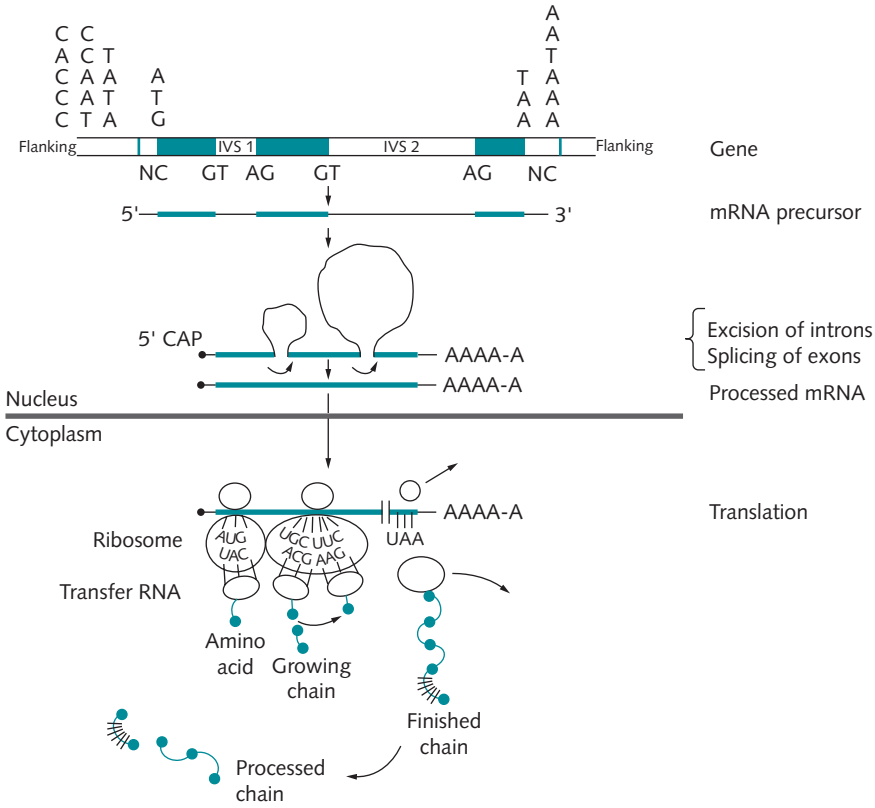

Fig. 1.2 The mechanisms of globin gene transcription and translation

the major functional regions of the particular genes. However, there are also extremely important regulatory sequences which subserve these functions that lie outside the genes themselves.

At the 5′ non-coding (flanking) regions of the globin genes, as in all mammalian genes, there are blocks of nucleotide homology. The first, the ATA box, is about 30 bases upstream (to the left) of the initiation codon; that is, the start word for the beginning of protein synthesis (*see below*). The second, the CCAAT box, is about 70 base pairs upstream from the 5′ end of the genes. About 80–100 bases further upstream there is the sequence GGGGTG, or CACCC, which may be inverted or duplicated. These three highly conserved DNA sequences, called *promoter elements*, are involved in the initiation of transcription of the individual genes. Finally, in the 3′ non-coding region of all the globin genes there is the sequence AATAAA, which is the signal for cleavage and polyA addition to RNA transcripts (*see section Gene action and globin synthesis*).

The globin gene clusters also contain several sequences that constitute regulatory elements, which interact to promote erythroid-specific gene expression and coordination of the changes in globin gene activity during development. These include the globin genes themselves and their promoter elements: enhancers (regulatory sequences that increase gene expression despite being located at a considerable distance from the genes) and "master" regulatory sequences called, in the case of the β globin gene cluster, the *locus control region* (LCR), and, in the case of the α genes, HS40 (a nuclease-hypersensitive site in DNA 40 kb from the α globin genes). Each of these sequences has a modular structure made up of an array of short motifs that represent the binding sites for transcriptional activators or repressors.

Gene action and globin synthesis

The flow of information between DNA and protein is summarized in Figure 1.2. When a globin gene is transcribed, messenger RNA (mRNA) is synthesized from one of its strands, a process which begins with the formation of a transcription complex consisting of a variety of regulatory proteins together with an enzyme called RNA polymerase (*see below*). The primary transcript is a large mRNA precursor which contains both intron and exon sequences. While in the nucleus, this molecule undergoes a variety of modifications. First, the introns are removed and the exons are spliced together. The intron/exon junctions always have the same sequence: GT at their 5′ end, and AG at their 3′ end. This appears to be essential for accurate splicing; if there is

a mutation at these sites this process does not occur. Splicing reflects a complex series of intermediary stages and the interaction of a number of different nuclear proteins. After the exons are joined, the mRNAs are modified and stabilized; at their 5′ end a complex CAP structure is formed, while at their 3′ end a string of adenylic acid residues (polyA) is added. The mRNA processed in this way moves into the cytoplasm, where it acts as a template for globin chain production. Because of the rules of base pairing, i.e., cytosine always pairs with thymine, and guanine with adenine, the structure of the mRNA reflects a faithful copy of the DNA codons from which it is synthesized; the only difference is that, in RNA, uracil (U) replaces thymine (T).

Amino acids are transported to the mRNA template on carriers called transfer RNAs (tRNAs); there are specific tRNAs for each amino acid. Furthermore, because the genetic code is redundant (i.e., more than one codon can encode a particular amino acid), for some of the amino acids there are several different individual tRNAs. Their order in the globin chain is determined by the order of codons in the mRNA. The tRNAs contain three bases, which together constitute an anticodon; these anticodons are complementary to mRNA codons for particular amino acids. They carry amino acids to the template, where they find the appropriate positioning by codon–anticodon base-pairing. When the first tRNA is in position, an initiation complex is formed between several protein initiation factors together with the two subunits that constitute the ribosomes. A second tRNA moves in alongside and the two amino acids that they are carrying form a peptide bond between them; the globin chain is now two amino acid residues long. This process is continued along the mRNA from left to right, and the growing peptide chain is transferred from one incoming tRNA to the next; that is, the mRNA is translated from 5′ to 3′. During this time the tRNAs are held in appropriate steric configuration with the mRNA by the two ribosomal subunits. There are specific initiation (AUG) and termination (UAA, UAG and UGA) codons. When the ribosomes reach the termination codon, translation ceases, the completed globin chains are released, and the ribosomal subunits are recycled. Individual globin chains combine with heme, which has been synthesized through a separate pathway, and then interact with one like chain and two unlike chains to form a complete hemoglobin tetramer.

Regulation of hemoglobin synthesis

The regulation of globin gene expression is mediated mainly at the transcriptional level, with some fine tuning during translation and post-translational modification of the gene products. DNA that is not involved in transcription is held tightly packaged in a compact, chemically modified form that is inaccessible to transcription factors and polymerases and which is heavily methylated. Activation of a particular gene is reflected by changes in the structure of the surrounding chromatin, which can be identified by enhanced sensitivity to nucleases. Erythroid lineage-specific nuclease-hypersensitive sites are found at several locations in the β globin gene cluster. Four are distributed over 20 kb upstream from the ε globin gene in the region of the β globin LCR (Figure 1.3). This vital regulatory region is able to establish a transcriptionally active domain spanning the entire β globin gene cluster. Several enhancer sequences have been identified in this cluster. A variety of regulatory proteins bind to the LCR, and to the promoter regions of the globin genes and to the enhancer sequences. It is thought that the LCR and other enhancer regions become opposed to the promoters to increase the rate of transcription of the genes to which they are related.

These regulatory regions contain sequence motifs for various ubiquitous and erythroid-restricted transcription factors. Binding sites for these factors have been identified in each of the globin gene promoters and at the hypersensitive-site regions of the various regulatory elements. A number of the factors which bind to these areas are found in all cell types. They include Sp1, Yy1 and Usf. In contrast, a number of transcription factors have been identified, including GATA-1, EKLF and NF-E2, which are restricted in their distribution to erythroid cells and, in some cases, megakaryocytes and mast cells. The overlapping of erythroid-specific and ubiquitous-factor binding sites in several cases suggests that competitive binding may play an important part in the regulation of erythroid-specific genes. Another binding factor, SSP, the stage selector protein, appears to interact specifically with ε and γ genes. Several elements involving the chromatin and histone acetylation required for access of these regulatory proteins have been identified.

The binding of hematopoietic-specific factors activates the LCR, which renders the entire β globin gene cluster transcriptionally active. These factors also bind to the enhancer and promoter sequences, which work in tandem to regulate the expression of the individual genes in the clusters. It is likely that some of the transcriptional factors are developmental stage-specific, and hence may be responsible for the differential expression of the embryonic, fetal and adult globin genes. The α globin gene cluster also contains an element, HS40, which has some structural features in common with the β LCR, although it is different in aspects of its structure. A number of enhancer-like sequences have also been identified, although it is becoming clear that there are fundamental differences in the pattern of regulation of the two globin gene clusters.

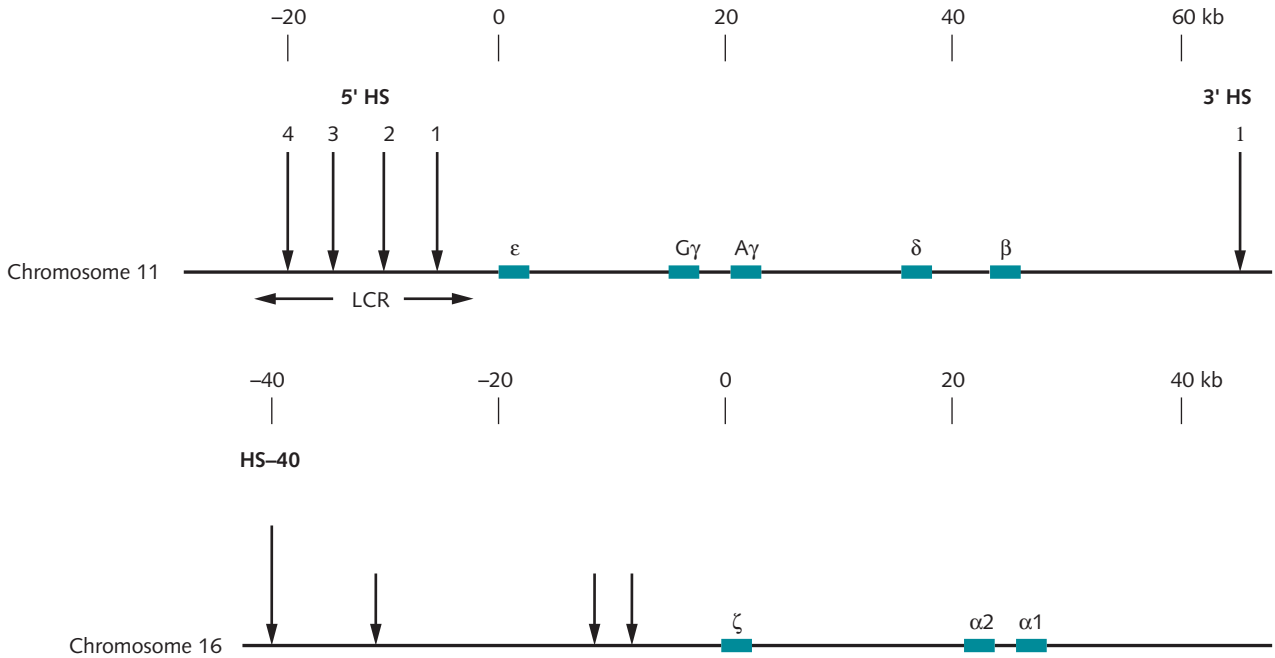

Fig. 1.3 The positions of the major regulatory regions in the β and α globin gene clusters
The arrows indicate the position of the erythroid lineage-specific nuclease-hypersensitive sites. HS, hypersensitive.

In addition to the different regulatory sequences outlined above, there are also sequences which may be involved specifically with "silencing" of genes, notably those for the embryonic hemoglobins, during development.

Some degree of regulation is also mediated by differences in the rates of initiation and translation of the different mRNAs, and at the post-transcriptional level by differential affinity for different protein subunits. However, this kind of post-transcriptional fine tuning probably plays a relatively small role in determining the overall output of the globin gene products.

Regulation of developmental changes in globin gene expression

During development, the site of red cell production moves from the yolk sac to the fetal liver and spleen, and thence to bone marrow in the adult. Embryonic, fetal and adult hemoglobin synthesis is approximately related in time to these changes in the site of erythropoiesis, although it is quite clear that the various switches, between embryonic and fetal and between fetal and adult hemoglobin synthesis, are beautifully synchronized throughout these different sites. Fetal hemoglobin synthesis declines during the later months of gestation and Hb F is replaced by Hb A and Hb A_2 by the end of the first year of life.

Despite a great deal of research, very little is known about the regulation of these different switches from one globin gene to another during development. Work from a variety of different sources suggests that there may be specific regions in the α and β globin gene clusters that are responsive to the action of transcription factors, some of which may be developmental-stage-specific. However, proteins of this type have not yet been isolated, and nothing is known about their regulation and how it is mediated during development.

The molecular pathology of hemoglobin

As is the case for most monogenic diseases, the inherited disorders of hemoglobin fall into two major classes. First, there are those that result from reduced output of one or other globin genes, the *thalassemias*. Second, there is a wide range of conditions that result from the production of *structurally abnormal globin chains*; the type of disease depends on how the particular alteration in protein structure interferes with its stability or function. Of course, no biological classification is entirely satisfactory and those which attempt to define the hemoglobin disorders are no exception. There are some structural hemoglobin variants which happen to be synthesized at a reduced rate and hence are associated with a clinical picture similar to thalassemia. And there are other classes of mutations which simply interfere with the normal transition from fetal to adult hemoglobin synthesis,

a family of conditions given the general title *hereditary persistence of fetal hemoglobin*. Furthermore, because these diseases are all so common and occur together in particular populations, it is not uncommon for an individual to inherit a gene for one or other form of thalassemia and a structural hemoglobin variant. The heterogeneous group of conditions that results from these different mutations and interactions is summarized in Table 1.1.

Over recent years, determination of the molecular pathology of the two common forms of thalassemia, α and β, has provided a remarkable picture of the repertoire of mutations that can underlie human monogenic disease. In the sections that follow I describe, in outline, the different forms of molecular pathology that underlie these conditions.

Table 1.1 The thalassemias and related disorders.

α Thalassemia	**γ Thalassemia**
α^0	
α^+	**δ Thalassemia**
Deletion (–α)	
Non-deletion (α^T)	**εγδβ Thalassemia**
β Thalassemia	**Hereditary persistence of fetal hemoglobin**
β^0	Deletion
β^+	$(\delta\beta)^0$
Normal Hb A_2	Non-deletion
"Silent"	Linked to β globin genes
	$^G\gamma\beta^+$
δβ Thalassemia	$^A\gamma\beta^+$
$(\delta\beta)^+$	Unlinked to β globin genes
$(\delta\beta)^0$	
$(^A\gamma\delta\beta)^0$	

The β thalassemias

There are two main classes of β thalassemia, β^0 thalassemia, in which there is an absence of β globin chain production, and β^+ thalassemia, in which there is variable reduction in the output of β globin chains. As shown in Figure 1.4, mutations of the β globin genes may cause a reduced output of gene product at the level of transcription or mRNA processing, translation, or through the stability of the globin gene product.

Defective β globin gene transcription

There are a variety of mechanisms that interfere with normal transcription of the β globin genes. First, the genes may be either completely or partially deleted. Overall, deletions of the β globin genes are not commonly found in patients with β thalassemia, with one exception: a 619-bp deletion involving the 3′ end of the gene is found frequently in the Sind populations of India and Pakistan, where it constitutes about 30% of the β thalassemia alleles. Other deletions are extremely rare.

A much more common group of mutations, which results in a moderate decrease in the rate of transcription of the β globin genes, involves single nucleotide substitutions in or near the TATA box at about −30 nucleotides (nt) from the transcription start site, or in the proximal or distal promoter elements at −90 nt and −105 nt. These mutations result in decreased β globin mRNA production, ranging from 10 to 25% of the normal output. Thus, they are usually associated with the mild forms of β^+ thalassemia. They are particularly common in African populations, an observation which explains the unusual mildness of β thalassemia in this racial

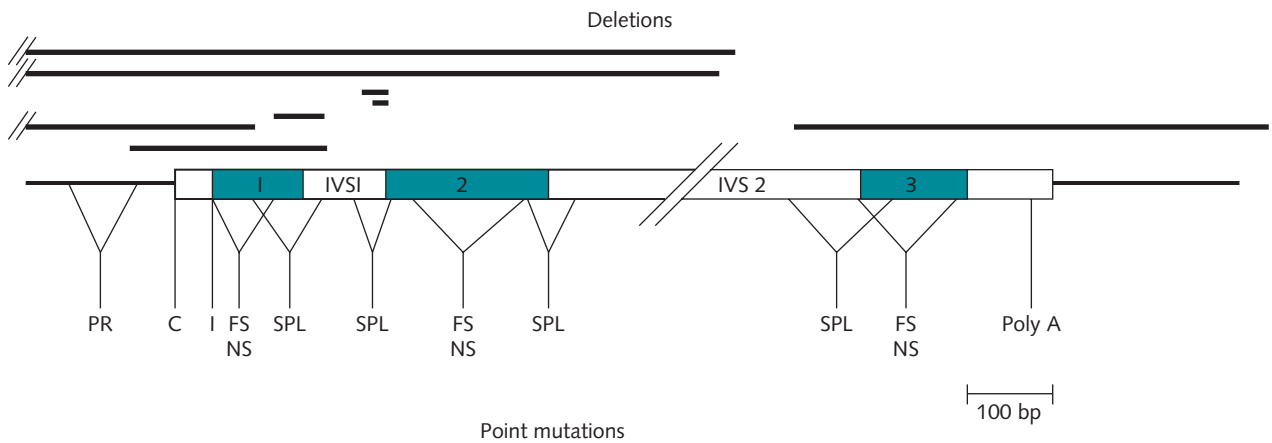

Fig. 1.4 The mutations of the β globin gene that underlie β thalassemia
The heavy black lines indicate the length of the deletions. The point mutations are designated as follows: PR, promoter; C, CAP site; I, initiation codon; FS, frameshift and nonsense mutations; SPL, splice mutations; Poly A, poly A addition site mutations.

group. One particular mutation, C→T at position −101 nt to the β globin gene, causes an extremely mild deficit of β globin mRNA. Indeed, this allele is so mild that it is completely silent in carriers and can only be identified by its interaction with more severe β thalassemia alleles in compound heterozygotes.

Mutations that cause abnormal processing of mRNA

As mentioned earlier, the boundaries between exons and introns are marked by the invariant dinucleotides GT at the donor (5′) site and AG at the acceptor (3′) site. Mutations that affect either of these sites completely abolish normal splicing and produce the phenotype of β^0 thalassemia. The transcription of genes carrying these mutations appears to be normal, but there is complete inactivation of splicing at the altered junction.

Another family of mutations involves what are called *splice site consensus sequences*. Although only the GT dinucleotide is invariant at the donor splice site, there is conservation of adjacent nucleotides and a common, or consensus, sequence of these regions can be identified. Mutations within this sequence can reduce the efficiency of splicing to varying degrees because they lead to alternate splicing at the surrounding cryptic sites. For example, mutations of the nucleotide at position 5 of IVS-1 (the first intervening sequence), G→C or T, result in a marked reduction of β chain production and in the phenotype of severe β^+ thalassemia. On the other hand, the substitution of C for T at position 6 in IVS-1 leads to only a mild reduction in the output of β chains.

Another mechanism that leads to abnormal splicing involves *cryptic splice sites*. These are regions of DNA which, if mutated, assume the function of a splice site at an inappropriate region of the mRNA precursor. For example, a variety of mutations activate a cryptic site which spans codons 24–27 of exon 1 of the β globin gene. This site contains a GT dinucleotide, and adjacent substitutions that alter it so that it more closely resembles the consensus donor splice site result in its activation, even though the normal splice site is intact. A mutation at codon 24 GGT→GGA, though it does not alter the amino acid which is normally found in this position in the β globin chain (glycine), allows some splicing to occur at this site instead of the exon–intron boundary. This results in the production of both normal and abnormally spliced β globin mRNA and hence in the clinical phenotype of severe β thalassemia. Interestingly, mutations at codons 19, 26 and 27 result in both reduced production of normal mRNA (due to abnormal splicing) and an amino acid substitution when the mRNA which is spliced normally is translated into protein. The abnormal hemoglobins produced are Hb Malay, Hb E and Hb Knossos, respectively. All these variants are associated with a mild β^+ thalassemia-like phenotype. These mutations illustrate how sequence changes in coding rather than intervening sequences influence RNA processing, and underline the importance of competition between potential splice site sequences in generating both normal and abnormal varieties of β globin mRNA.

Cryptic splice sites in introns may also carry mutations that activate them even though the normal splice sites remain intact. A common mutation of this kind in Mediterranean populations involves a base substitution at position 110 in IVS-1. This region contains a sequence similar to a 3′ acceptor site, though it lacks the invariant AG dinucleotide. The change of the G to A at position 110 creates this dinucleotide. The result is that about 90% of the RNA transcript splices to this particular site and only 10% to the normal site, again producing the phenotype of severe β^+ thalassemia (Figure 1.5). Several other β thalassemia mutations have been described which generate new donor sites within IVS-2 of the β globin gene.

Another family of mutations that interferes with β globin gene processing involves the sequence AAUAAA in the 3′ untranslated regions, which is the signal for cleavage and polyadenylation of the β globin gene transcript. Somehow, these mutations destabilize the transcript. For example, a T→C substitution in this sequence leads to only one-tenth of the normal amount of β globin mRNA transcript and hence to the phenotype of a moderately severe β^+ thalassemia. Another example of a mutation which probably leads to defective processing of function of β globin mRNA is the single-base substitution A→C in the CAP site. It is not yet understood how this mutation causes a reduced rate of transcription of the β globin gene.

There is another small subset of rare mutations that involve the 3′ untranslated region of the β globin gene and these are associated with relatively mild forms of β thalassemia. It is thought that these interfere in some way with transcription but the mechanism is unknown.

Mutations that result in abnormal translation of β globin mRNA

There are three main classes of mutations of this kind. Base substitutions that change an amino acid codon to a chain termination codon prevent the translation of β globin mRNA and result in the phenotype of β^0 thalassemia. Several mutations of this kind have been described; the commonest, involving codon 17, occurs widely throughout Southeast Asia. Similarly, a codon 39 mutation is encountered frequently in the Mediterranean region.

The second class involves the insertion or deletion of one, two or four nucleotides in the coding region of the β globin gene. These disrupt the normal reading frame, cause a

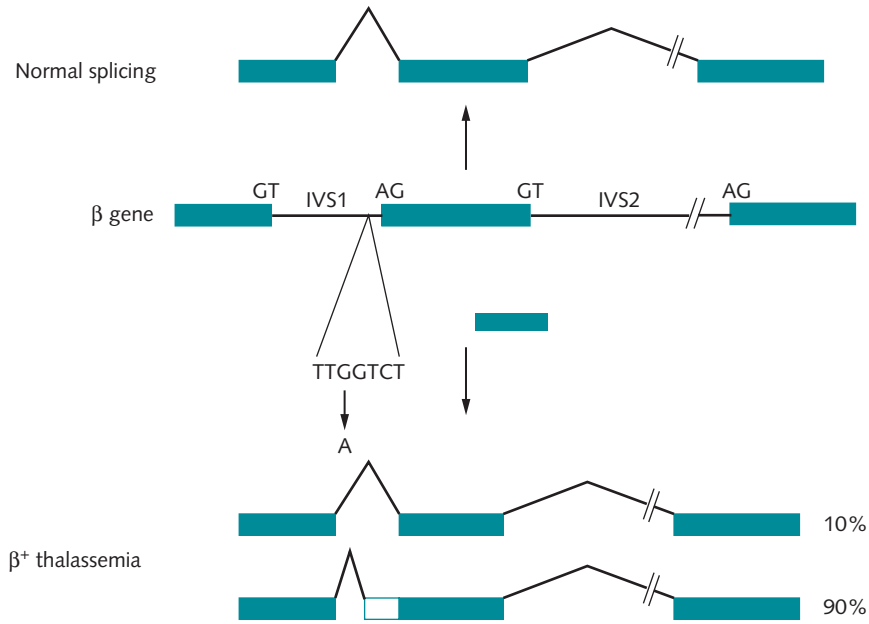

Fig. 1.5 The generation of a new splice site in an intron as the mechanism for a form of β^+ thalassemia
For details see text.

frameshift, and hence interfere with the translation of β globin mRNA. The end result is the insertion of anomalous amino acids after the frameshift until a termination codon is reached in the new reading frame. This type of mutation always leads to the phenotype of β^0 thalassemia.

Finally, there are several mutations which involve the β globin gene initiation codon and which, presumably, reduce the efficiency of translation.

Unstable β globin chain variants

Some forms of β thalassemia result from the synthesis of highly unstable β globin chains that are incapable of forming hemoglobin tetramers, and which are rapidly degraded, leading to the phenotype of β^0 thalassemia. Indeed, in many of these conditions no abnormal globin chain product can be demonstrated by protein analysis and the molecular pathology has to be interpreted simply on the basis of a derived sequence of the variant β chain obtained by DNA analysis.

Recent studies have provided some interesting insights into how complex clinical phenotypes may result from the synthesis of unstable β globin products. For example, there is a spectrum of disorders that result from mutations in exon 3 which give rise to a moderately severe form of β thalassemia in heterozygotes. It has been found that nonsense or frameshift mutations in exons I and II are associated with the absence of mRNA from the cytoplasm of red cell precursors. This appears to be an adaptive mechanism, called *nonsense-mediated decay*, whereby abnormal mRNA of this type is not transported to the cytoplasm, where it would act as a template for the production of truncated gene products. However, in the case of exon III mutations, apparently because this process requires the presence of an intact upstream exon, the abnormal mRNA is transported into the cytoplasm and hence can act as a template for the production of unstable β globin chains. The latter precipitate in the red cell precursors together with excess α chains to form large inclusion bodies, and hence there is enough globin chain imbalance in heterozygotes to produce a moderately severe degree of anemia.

The α thalassemias

The molecular pathology of the α thalassemias is more complicated than that of the β thalassemias, simply because there are two α globin genes per haploid genome. Thus, the normal α globin genotype can be written αα/αα. As in the case of β thalassemia, there are two major varieties of α thalassemia, α^+ and α^0 thalassemia. In α^+ thalassemia one of the linked α globin genes is lost, either by deletion (–) or mutation (T); the heterozygous genotype can be written –α/αα or $\alpha^T\alpha/\alpha\alpha$. In α^0 thalassemia the loss of both α globin genes nearly always results from a deletion; the heterozygous genotype is therefore written ––/αα. In populations where specific deletions are particularly common, Southeast Asia (SEA) or the Mediterranean region (MED), it is useful to add the appropriate superscript as follows: $--^{SEA}/\alpha\alpha$ or $--^{MED}/\alpha\alpha$. It follows that when we speak of an "α thalassemia gene" what we are really referring to is a haplotype; that is, the state and function of both of the linked α globin genes.

α^0 Thalassemia

Three main molecular pathologies, all involving deletions, have been found to underlie the α^0 thalassemia phenotype. The majority of cases result from deletions that remove both α globin genes and a varying length of the α globin gene cluster (Figure 1.6). Occasionally, however, the α globin gene cluster is intact but is inactivated by a deletion which involves the major regulatory region HS40, 40 kb upstream from the α globin genes, or the α globin genes may be lost as part of a truncation of the tip of the short arm of chromosome 16.

As well as providing us with an understanding of the molecular basis for α^0 thalassemia, detailed studies of these deletions have yielded more general information about the mechanisms that underlie this form of molecular pathology. For example, it has been found that the 5′ breakpoints of a number of deletions of the α globin gene cluster are located approximately the same distance apart and in the same order along the chromosome as their respective 3′ breakpoints; similar findings have been observed in deletions of the β globin gene cluster. These deletions seem to have resulted from illegitimate recombination events which have led to the deletion of an integral number of chromatin loops as they pass through their nuclear attachment points during chromosomal replication. Another long deletion has been characterized in which a new piece of DNA bridges the two breakpoints in the α globin gene cluster. The inserted sequence originates upstream from the α globin gene cluster, where normally it is found in an inverted orientation with respect to that found between the breakpoints of the deletion. Thus it appears to have been incorporated into the junction in a way that reflects its close proximity to the deletion breakpoint region during replication. Other deletions seem to be related to the family of Alu-repeats, simple repeat sequences that are widely dispersed throughout the genome; one deletion appears to have resulted from a simple homologous recombination between two repeats of this kind that are usually 62 kb apart.

A number of forms of α^0 thalassemia result from terminal truncations of the short arm of chromosome 16 to a site about 50 kb distal to the α globin genes. The telomeric consensus sequence $TTAGGG_n$ has been added directly to the site of the break. Since these mutations are stably inherited, it appears that telomeric DNA alone is sufficient to stabilize the ends of broken chromosomes.

Quite recently, two other molecular mechanisms have been identified as the cause of α^0 thalassemia which, though rare, may have important implications for an understanding of the molecular pathology of other genetic diseases. In one case, a deletion in the α globin gene cluster resulted in a widely expressed gene (*LUC7L*) becoming juxtaposed to a structurally normal α globin gene. Although the latter retained all its important regulatory elements, its expression was silenced. It was found in a transgenic mouse model that transcription of antisense RNA mediated the silencing of the α globin gene region, findings that provide a completely new mechanism for genetic disease. In another case of α^0 thalassemia, in which no molecular defects could be detected in the α globin gene cluster, a gain-of-function regulatory polymorphism was found in the region between the α globin genes and their upstream regulatory elements. This alteration creates a new promoter-like element that interferes with the normal activation of all downstream α-like globin genes.

In short, detailed analysis of the molecular pathology of the α^0 thalassemias has provided valuable evidence not only about how large deletions of gene clusters are caused, but also about some of the complex mechanisms that may underlie cases in which the α gene clusters remain intact but in which their function is completely suppressed.

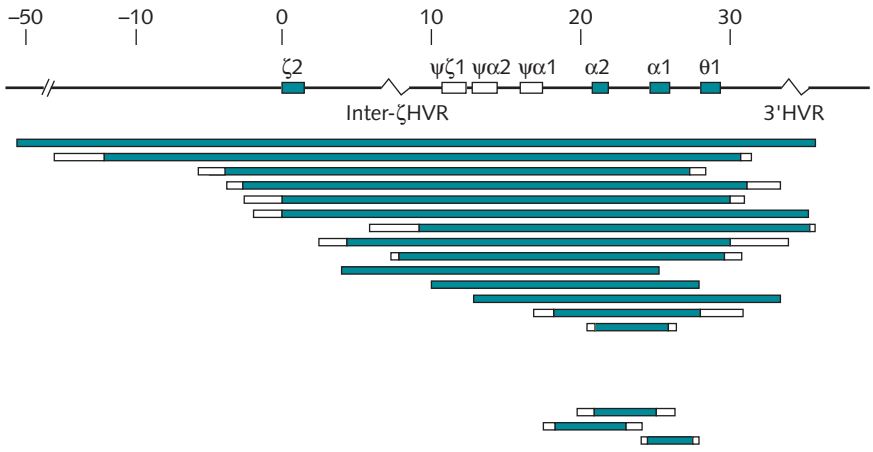

Fig. 1.6 Some of the deletions that underlie α^0 and α^+ thalassemia
The colored rectangles beneath the α globin gene cluster indicate the lengths of the deletions. The unshaded regions indicate uncertainty about the precise breakpoints. The three small deletions at the bottom of the figure represent the common α^+ thalassemia deletions. HVR, highly variable regions.

α+ Thalassemia

As mentioned earlier, the α+ thalassemias result from the inactivation of one of the duplicated α globin genes, either by deletion or point mutation.

α+ Thalassemia due to gene deletions There are two common forms of α+ thalassemia that are due to loss of one or other of the duplicated α globin genes, $-\alpha^{3.7}$ and $-\alpha^{4.2}$, where 3.7 and 4.2 indicate the size of the deletions. The way in which these deletions have been generated reflects the underlying structure of the α globin gene complex (Figure 1.7). Each α gene lies within a boundary of homology, approximately 4 kb long, probably generated by an ancient duplication event. The homologous regions, which are divided by small inserts, are designated X, Y and Z. The duplicated Z boxes are 3.7 kb apart and the X boxes are 4.2 kb apart. As the result of misalignment and reciprocal crossover between these segments at meiosis, a chromosome is produced with either a single (–α) or triplicated (ααα) α globin gene. As shown in Figure 1.7, if a crossover occurs between homologous Z boxes 3.7 kb of DNA are lost, an event which is described as a rightward deletion, $-\alpha^{3.7}$. A similar crossover between the two X boxes deletes 4.2 kb, the leftward deletion $-\alpha^{4.2}$. The corresponding triplicated α gene arrangements are called $\alpha\alpha\alpha^{anti\,3.7}$ and $\alpha\alpha\alpha^{anti\,4.2}$. A variety of different points of crossing over within the Z boxes give rise to different length deletions, still involving 3.7 kb.

Non-deletion types of α+ thalassemia These disorders result from single or oligonucleotide mutations of the particular α globin gene. Most of them involve the α2 gene but, since the output from this locus is two to three times greater than that from the α1 gene, this may simply reflect ascertainment bias due to the greater phenotypic effect and, possibly, a greater selective advantage.

Overall, these mutations interfere with α globin gene function in a similar way to those that affect the β globin genes. They affect the transcription, translation or post-translational stability of the gene product. Since the principles are the same as for β thalassemia, we do not need to describe them in detail with one exception, a mutation which has not been observed in the β globin gene cluster. It turns out that there is a family of mutations that involves the α2 globin gene termination codon, TAA. Each specifically changes this codon so that an amino acid is inserted instead of the chain terminating. This is followed by "read-through" of α globin mRNA, which is not normally translated until another in-phase termination codon is reached. The result is an elongated α chain with 31 additional residues at the C-terminal end. Five hemoglobin variants of this type have been identified. The commonest, Hb Constant Spring, occurs at a high frequency in many parts of Southeast Asia. It is not absolutely clear why the read-through of normally untranslated mRNAs leads to a reduced output from the α2 gene, although there is considerable evidence that it in some way destabilizes the mRNA.

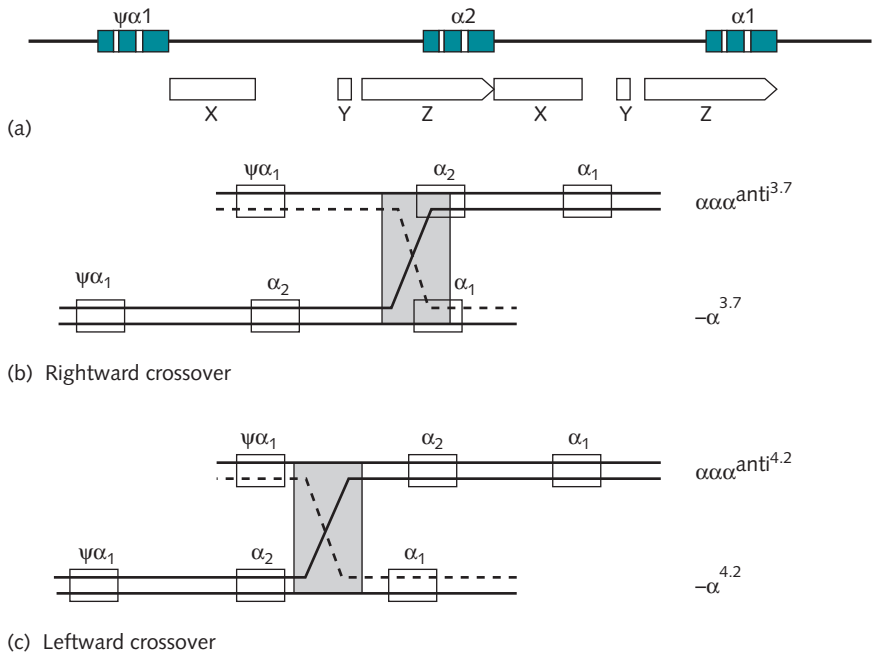

Fig. 1.7 Mechanisms of the generation of the common deletion forms of α+ thalassemia
(a) The normal arrangement of the α globin genes, with the regions of homology X, Y and Z. (b) The crossover that generates the $-\alpha^{3.7}$ deletion. (c) The crossover that generates the $-\alpha^{4.2}$ deletion.

α Thalassemia/mental retardation syndromes

There is a family of mild forms of α thalassemia which is quite different to that described in the previous section and which is associated with varying degrees of mental retardation. Recent studies indicate that there are two quite different varieties of this condition, one encoded on chromosome 16 (ATR-16) and the other on the X chromosome (ATR-X).

The ATR-16 syndrome is characterized by relatively mild mental handicap with a variable constellation of facial and skeletal dysmorphisms. These individuals have long deletions involving the α globin gene cluster, but removing at least 1–2 Mb. This condition can arise in several ways, including unbalanced translocation involving chromosome 16, truncation of the tip of chromosome 16, and the loss of the α globin gene cluster and parts of its flanking regions by other mechanisms.

The ATR-X syndrome results from mutations in a gene on the X chromosome, Xq13.1–q21.1. The product of this gene is one of a family of proteins involved in chromatin-mediated transcriptional regulation. It is expressed ubiquitously during development and at interphase it is found entirely within the nucleus in association with pericentromeric heterochromatin. In metaphase, it is similarly found close to the centromeres of many chromosomes but, in addition, occurs at the stalks of acrocentric chromosomes, where the sequences for ribosomal RNA are located. These locations provide important clues to the potential role of this protein in the establishment and/or maintenance of methylation of the genome. Although it is clear that *ATR-X* is involved in α globin transcription, it also must be an important player in early fetal development, particularly of the urogenital system and brain. Many different mutations of this gene have been discovered in association with the widespread morphological and developmental abnormalities which characterize the ATR-X syndrome.

α Thalassemia and the myelodysplastic syndrome

Since the first description of Hb H (*see later section*) in the red cells of a patient with leukemia, many examples of this association have been reported. The condition usually is reflected in a mild form of Hb H disease, with typical Hb H inclusions in a proportion of the red cells and varying amounts of Hb H demonstrable by hemoglobin electrophoresis. The hematological findings are usually those of one or other form of the myelodysplastic syndrome. The condition occurs predominantly in males in older age groups. Very recently it has been found that some patients with this condition have mutations involving *ATR-X*. The relationship of these mutations to the associated myelodysplasia remains to be determined.

Rarer forms of thalassemia and related disorders

There are a variety of other conditions that involve the β globin gene cluster which, although less common than the β thalassemias, provide some important information about mechanisms of molecular pathology and therefore should be mentioned briefly.

The δβ thalassemias

Like the β thalassemias, the δβ thalassemias, which result from defective δ and β chain synthesis, are subdivided into the $(\delta\beta)^+$ and $(\delta\beta)^0$ forms.

The $(\delta\beta)^+$ thalassemias result from unequal crossover between the δ and β globin gene loci at meiosis with the production of δβ fusion genes. The resulting δβ fusion chain products combine with α chains to form a family of hemoglobin variants called the hemoglobin Lepores, after the family name of the first patient of this kind to be discovered. Because the synthesis of these variants is directed by genes with the 5′ sequences of the δ globin genes, which have defective promoters, they are synthesized at a reduced rate and result in the phenotype of a moderately severe form of δβ thalassemia.

The $(\delta\beta)^0$ thalassemias nearly all result from long deletions involving the β globin gene complex. Sometimes they involve the $^{A}\gamma$ globin chains and hence the only active locus remaining is the $^{G}\gamma$ locus. In other cases the $^{G}\gamma$ and $^{A}\gamma$ loci are left intact and the deletion simply removes the δ and β globin genes; in these cases both the $^{G}\gamma$ and the $^{A}\gamma$ globin gene remains functional. For some reason, these long deletions allow persistent synthesis of the γ globin genes at a relatively high level during adult life, which helps to compensate for the absence of β and δ globin chain production. They are classified according to the kind of fetal hemoglobin that is produced, and hence into two varieties, $^{G}\gamma(^{A}\gamma\delta\beta)^0$ and $^{G}\gamma^{A}\gamma(\delta\beta)^0$ thalassemia; in line with other forms of thalassemia, they are best described by what is not produced: $(^{A}\gamma\delta\beta)^0$ and $(\delta\beta)^0$ thalassemia, respectively. Homozygotes produce only fetal hemoglobin, while heterozygotes have a thalassemic blood picture together with about 5–15% Hb F.

Hereditary persistence of fetal hemoglobin

Genetically determined persistent fetal hemoglobin synthesis in adult life is of no clinical importance except that its genetic determinants can interact with the β thalassemias or

structural hemoglobin variants; the resulting high level of Hb F production often ameliorates these conditions. The different forms of hereditary persistence of fetal hemoglobin (HPFH) result from either long deletions involving the δβ globin gene cluster, similar to those that cause $(\delta\beta)^0$ thalassemia, or from point mutations that involve the promoters of the $^{G}\gamma$ or $^{A}\gamma$ globin gene. In the former case there is no β globin chain synthesis and therefore these conditions are classified as $(\delta\beta)^0$ HPFH. In cases in which there are promoter mutations involving the γ globin genes, there is increased γ globin chain production in adult life associated with some β and δ chain synthesis in *cis* (i.e., directed by the same chromosome) to the HPFH mutations. Thus, depending on whether the point mutations involve the promoter of the $^{G}\gamma$ or $^{A}\gamma$ globin gene, these conditions are called $^{G}\gamma\ \beta^+$ HPFH and $^{A}\gamma\ \beta^+$ HPFH, respectively.

There is another family of HPFH-like disorders in which the genetic determinant is not encoded in the β chain cluster. In one case the determinant encodes on chromosome 6, although its nature has not yet been determined.

It should be pointed out that all these conditions are very heterogeneous and that many different deletions or point mutations have been discovered that produce the rather similar phenotypes of $(\delta\beta)^0$ or $^{G}\gamma$ or $^{A}\gamma\ \beta^+$ HPFH.

Genotype–phenotype relationships in the thalassemias

It is now necessary briefly to relate the remarkably diverse molecular pathology described in the previous sections to the phenotypes observed in patients with these diseases. It is not possible to describe all these complex issues here. Rather we shall focus on those aspects that illustrate the more general principles of how abnormal gene action is reflected in a particular clinical picture. Perhaps the most important question that we will address is why patients with apparently identical genetic lesions have widely differing disorders, a problem that still bedevils the whole field of medical genetics, even in the molecular era.

The β thalassemias

As we have seen, the basic defect that results from the 200 or more different mutations that underlie these conditions is reduced β globin chain production. Synthesis of the α globin chain proceeds normally and hence there is imbalanced globin chain output with an excess of α chains (Figure 1.8). Unpaired α chains precipitate in both red cell precursors and their progeny with the production of inclusion bodies. These interfere with normal red cell maturation and survival in a variety of complex ways. Their attachment to the red cell membrane causes alterations in its structure, and their degradation products, notably heme, hemin (oxidized heme) and iron, result in oxidative damage to the red cell contents and membrane. These interactions result in intramedullary destruction of red cell precursors and in shortened survival of such cells as they reach the peripheral blood. The end result is an anemia of varying severity. This, in turn, causes tissue hypoxia and the production of relatively large amounts of erythropoietin; this leads to a massive expansion of the ineffective bone marrow, resulting in bone deformity, a hypermetabolic state with wasting and malaise, and bone fragility.

A large proportion of hemoglobin in the blood of β thalassemics is of the fetal variety. Normal individuals produce about 1% of Hb F, unevenly distributed among their red cells. In the bone marrow of β thalassemics, any red cell precursors that synthesize γ chains come under strong selection because they combine with α chains to produce fetal hemoglobin and therefore the degree of globin chain imbalance is reduced. Furthermore, the likelihood of γ chain production seems to be increased in a highly stimulated erythroid bone marrow. It seems likely that these two factors combine to increase the relative output of Hb F in this disorder. However, it has a higher oxygen affinity than Hb A and hence patients with β thalassemia are not able to adapt to low hemoglobin levels as well as those who have adult hemoglobin.

The greatly expanded, ineffective erythron leads to an increased rate of iron absorption; this, combined with iron received by blood transfusion, leads to progressive iron loading of the tissues, with subsequent liver, cardiac and endocrine damage.

The constant bombardment of the spleen with abnormal red cells leads to its hypertrophy. Hence there is progressive splenomegaly with an increased plasma volume and trapping of part of the circulating red cell mass in the spleen. This leads to worsening of the anemia. All these pathophysiological mechanisms, except for iron loading, can be reversed by regular blood transfusion which, in effect, shuts off the ineffective bone marrow and its consequences.

Thus it is possible to relate nearly all the important features of the severe forms of β thalassemia to the primary defect in globin gene action. However, can we also explain their remarkable clinical diversity?

Phenotypic diversity

Although the bulk of patients who are homozygous for β thalassemia mutations or compound heterozygotes for two different mutations have a severe transfusion-dependent

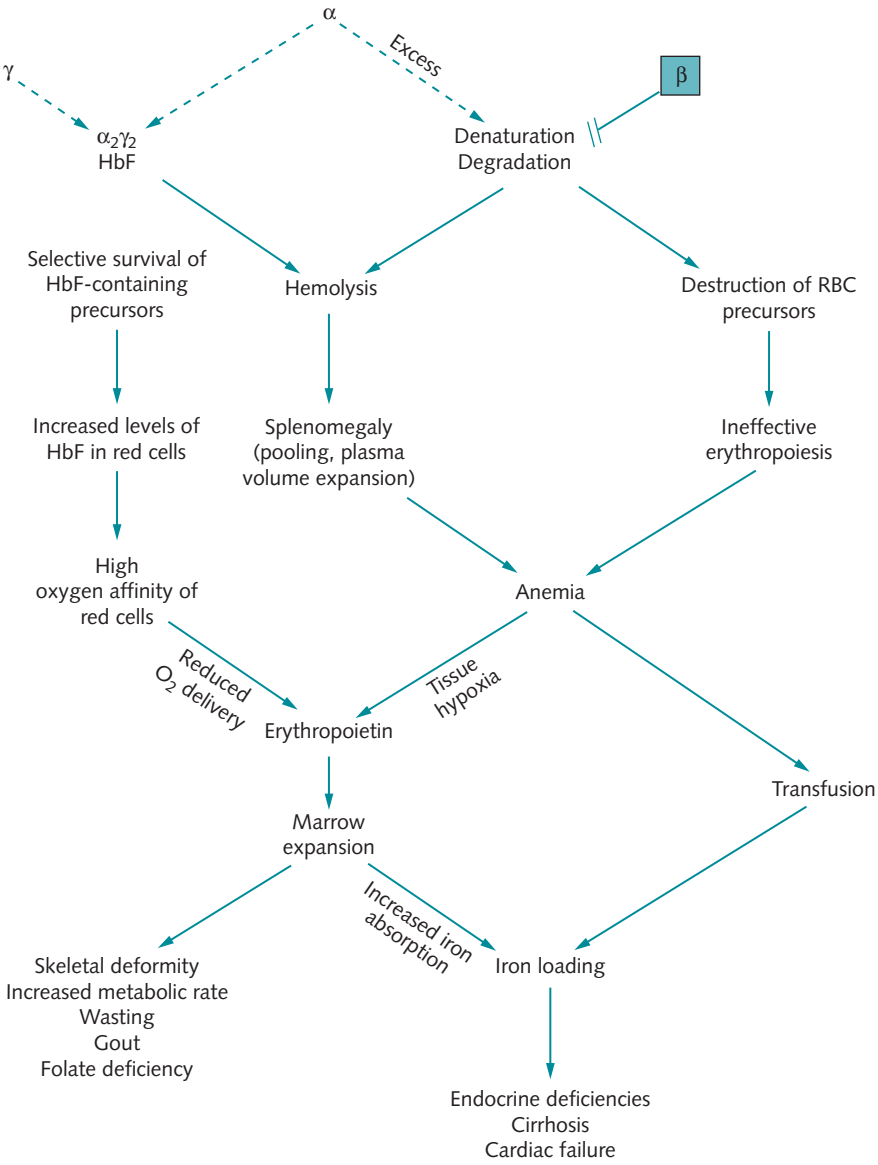

Fig. 1.8 The pathophysiology of β thalassemia

phenotype, there are many exceptions. Some patients of this type have a milder course, requiring few or even no transfusions, a condition called β thalassemia intermedia. A particularly important example of this condition is illustrated by the clinical findings in those who inherit β thalassemia from one parent and Hb E from the other, a disorder called Hb E/β thalassemia. Because the mutation that produces Hb E also opens up an alternative splice site in the first exon of the β globin gene, it is synthesized at a reduced rate and therefore behaves like a mild form of β thalassemia. It is the commonest hemoglobin variant globally and Hb E/β thalassemia is the commonest form of severe thalassemia in many Asian countries. It has an extraordinarily variable phenotype, ranging from a condition indistinguishable from β thalassemia major to one of such mildness that patients grow and develop quite normally and never require transfusion.

Over recent years a great deal has been learnt about some of the mechanisms involved in this remarkable phenotypic variability. In short, it reflects both the action of modifying genes and variability in adaptation to anemia and, almost certainly, the effects of the environment. Given the complexity of these interactions, it is helpful to divide the genetic modifiers of the β thalassemia phenotype into primary, secondary and tertiary classes (Table 1.2).

The primary modifiers are the different β thalassemia alleles that can interact together. For example, compound heterozygotes for a severe β^0 thalassemia mutation and a milder one may have an intermediate form of β thalassemia

Table 1.2 Mechanisms for the phenotypic diversity of the β thalassemias.

Genetic modifiers
Primary: alleles of varying severity
Secondary: modifiers of globin chain imbalance
α Thalassemia
Increased α globin genes: ααα or αααα
Genes involved in unusually high Hb F response
Tertiary: modifiers of complications
Iron absorption, bone disease, jaundice, infection
Adaptation to anemia*
Variation in oxygen affinity (P_{50}) of hemoglobin
Variation in erythropoietin response to anemia
Environmental
Nutrition
Infection
Others

*There may be genetic variation in the adaptive mechanisms.

of varying severity depending on the degree of reduction in β globin synthesis under the action of the milder allele. This is undoubtedly one mechanism for the varying severity of Hb E/β thalassemia; it simply reflects the variable action of the β thalassemia mutation that is inherited together with Hb E. However, this explanation is not relevant in cases in which patients with identical β thalassemia mutations have widely disparate phenotypes.

The secondary modifiers are those which directly affect the degree of globin chain imbalance. Patients with β thalassemia who also inherit one or other form of α thalassemia tend to have a milder phenotype because of the reduction in the excess of α globin genes caused by the coexistent α thalassemia allele. Similarly, patients with severe forms of thalassemia who inherit more α genes than normal because their parents have triplicated or quadruplicated α gene arrangements tend to have more severe phenotypes. Other patients with severe thalassemia alleles appear to run a milder course because of a genetically determined ability to produce more γ globin chains and hence fetal hemoglobin, a mechanism that also results in a reduced degree of globin chain imbalance. It is now clear that several gene loci are involved in this mechanism; the best characterized is a polymorphism in the promoter region of the $^{G}\gamma$ globin gene that appears to increase the output from this locus under conditions of hemopoietic stress. However, there are clearly other genes involved in increasing the output of Hb F. Recent genome-wide linkage studies have shown clear evidence that there are determinants on chromosomes 6 and 8 and a particularly strong association has been found with *BCL11A*, a transcription factor known to be involved in hematopoiesis. The exact mechanism for the associated increase in Hb F in β thalassemia and in sickle cell anemia remains to be determined.

The tertiary modifiers are those that have no effect on hemoglobin synthesis but which modify the many different complications of the β thalassemias, including osteoporosis, iron absorption, jaundice, and susceptibility to infection.

Although neglected until recently, it is also becoming apparent that variation in adaptation to anemia and the environment may also play a role in phenotypic modification of the β thalassemias. For example, patients with Hb E/β thalassemia have relatively low levels of Hb F and hence their oxygen dissociation curves are more right-shifted than patients with other forms of β thalassemia intermedia with significantly higher levels of Hb F. Very recent studies also suggest that the erythropoietin response to severe anemia for a given hemoglobin level varies considerably with age; patients during the first years of life have significantly higher responses to the same hemoglobin level than those who are older. This observation may go some way to explaining the variation in phenotype at different ages that has been observed in children with Hb E/β thalassemia. Finally, it is clear that further studies are required to dissociate the effects on the phenotype of genetic modifiers and environmental factors.

Thus the phenotypic variability of the β thalassemias reflects several layers of complex interactions involving genetic modifiers together with variation in adaptation and, almost certainly, the environment. These complex interactions are summarized in Table 1.2.

The α thalassemias

The pathophysiology of the α thalassemias differs from that of the β thalassemias mainly because of the properties of the excess globin chains that are produced as a result of defective α chain synthesis. While the excess α chains produced in β thalassemia are unstable and precipitate, this is not the case in the α thalassemias, in which excess γ chains or β chains are able to form the soluble homotetramers γ_4 (Hb Bart's) and β_4 (Hb H) (Figure 1.9). Although these variants, particularly Hb H, are unstable and precipitate in older red cell populations, they remain soluble sufficiently long for the red cells to mature and develop relatively normally. Hence there is far less ineffective erythropoiesis in the α thalassemias and the main cause of the anemia is hemolysis associated with the precipitation of Hb H in older red cells. In addition, of course, there is a reduction in normal hemoglobin synthesis, which results in hypochromic, microcytic erythrocytes. Another important factor in the pathophysiology of the α

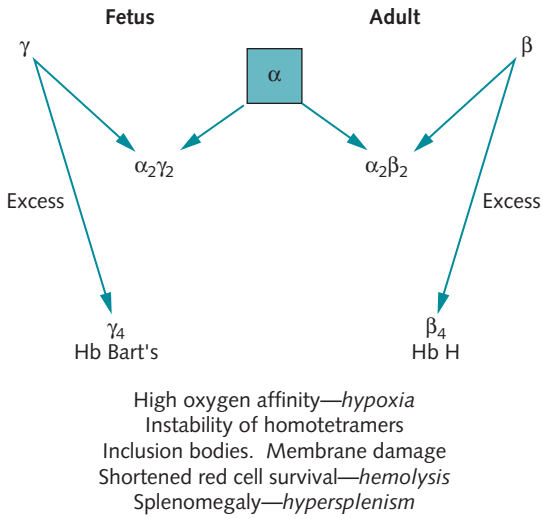

Fig. 1.9 The pathophysiology of α thalassemia

thalassemias is the fact that Hb Bart's and Hb H are useless oxygen carriers, having an oxygen dissociation curve similar to that of myoglobin. Hence the circulating hemoglobin level may give a false impression of the oxygen-delivering capacity of the blood and patients may be symptomatic at relatively high hemoglobin levels.

The different clinical phenotypes of the α thalassemias are an elegant example of the effects of gene dosage (Figure 1.10). The heterozygous state for α^+ thalassemia is associated with minimal hematological changes. That for α^0 thalassemia (the loss of two α globin genes) is characterized by moderate hypochromia and microcytosis, similar to that of the β thalassemia trait. It does not matter whether the α genes are lost on the same chromosome or on opposite pairs of homologous chromosomes. Hence the homozygous state for α^+ thalassemia, $-\alpha/-\alpha$, has a similar phenotype to the heterozygous state for α^0 thalassemia ($--/\alpha\alpha$).

The loss of three α globin genes, which usually results from the compound heterozygous states for α^0 and α^+ thalassemia, is associated with a moderately severe anemia with the production of varying levels of Hb H. This condition, hemoglobin H disease, is characterized by varying anemia and splenomegaly with a marked shortening of red cell survival.

Finally, the homozygous state for α^0 thalassemia ($--/--$) is characterized by death *in utero* or just after birth, with the clinical picture of hydrops fetalis. These babies produce no α chains and their hemoglobin consists mainly of Hb Bart's with variable persistence of embryonic hemoglobin. This is reflected in gross intrauterine hypoxia; although these babies may have hemoglobin values as high as 8–9 g/dL, most of it is unable to release its oxygen. This is reflected in the hydropic changes, a massive outpouring of nucleated red cells, and hepatosplenomegaly with persistent hematopoiesis in the liver and spleen.

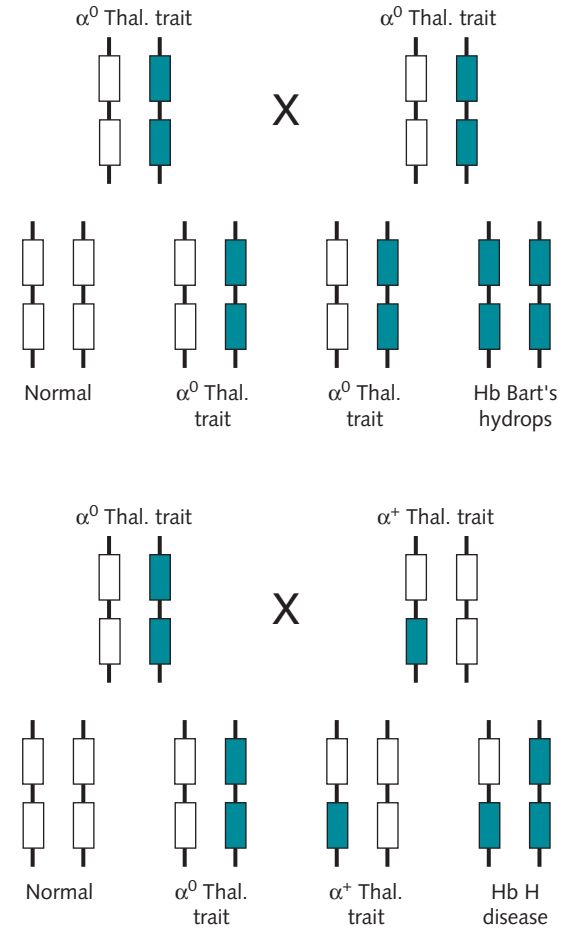

Fig. 1.10 The genetics of the common forms of α thalassemia
The open boxes represent normal α genes and the green boxes deleted α genes. The mating shown at the top shows how two α^0 thalassemia heterozygotes can produce a baby with the Hb Bart's hydrops syndrome. In the mating at the bottom, between individuals with α^0 and α^+ thalassemia, one in four of the offspring will have Hb H disease.

Structural hemoglobin variants

The structural hemoglobin variants are described in detail in *Chapter 15*. Here, their molecular pathology and genotype–phenotype relationships are briefly outlined.

Molecular pathology

The molecular pathology of the structural hemoglobin variants is much less complex than that of the thalassemias. The

majority result from missense mutations – base substitutions that produce a codon change which encodes a different amino acid in the affected globin chain. Rarely, these variants result from more subtle alterations in the structure of the α/β globin chains. For example, shortened chains may result from internal deletions of their particular genes, while elongated chains result from either duplications within genes or frameshift mutations which allow the chain termination codon to be read through and in which additional amino acids are added to the C-terminal end. The majority of the 700 or more structural hemoglobin variants are of no clinical significance but a few, because they interfere with the stability or functions of the hemoglobin molecule, are associated with a clinical phenotype of varying severity.

Genotype–phenotype relationships

The sickling disorders

The sickling disorders represent the homozygous state for the sickle cell gene, sickle cell anemia, and the compound heterozygous state for the sickle cell gene and various structural hemoglobin variants, or β thalassemia. The chronic hemolysis and episodes of vascular occlusion and red cell sequestration that characterize sickle cell anemia can all be related to the replacement of the normal β6 glutamic acid by valine in Hb S. This causes a hydrophobic interaction with another hemoglobin molecule, triggering aggregation into large polymers. It is this change that causes the sickling distortion of the red blood cell and hence a marked decrease in its deformability. The resulting rigidity of the red cells is responsible for the vaso-occlusive changes that lead to many of the most serious aspects of all the sickling disorders.

The different conformations of sickle cells (banana-shaped or resembling a holly leaf) reflect different orientations of bundles of fibers along the long axis of the cell, the three-dimensional structure of which is constituted by a rope-like polymer composed of 14 strands. The rate and extent of polymer formation depend on the degree of oxygenation, the cellular hemoglobin concentration, and the presence or absence of Hb F. The latter inhibits polymerization and hence tends to ameliorate sickling. Polymerization of Hb S causes damage to the red cell membrane, the result of which is an irreversibly sickled cell. Probably the most important mechanism is cellular dehydration resulting from abnormalities of potassium/chloride cotransport and Ca^{2+}-activated potassium efflux. This is sufficient to trigger the Ca^{2+}-dependent (Gardos) potassium channel, providing a mechanism for the loss of potassium and water and leading to cellular dehydration.

However, the vascular pathology of the sickling disorders is not entirely related to the rigidity of sickled red cells. There is now a wealth of evidence that abnormal interactions between sickled cells and the vascular endothelium play a major role in the pathophysiology of the sickling disorders. Recently it has been demonstrated that nitric oxide may also play a role in some of the vascular complications of this disease. It has been found that nitric oxide reacts much more rapidly with free hemoglobin than with hemoglobin in erythrocytes and therefore it is possible that such decompartmentalization of hemoglobin into plasma, as occurs in sickle cell disease and other hemolytic anemias, diverts nitric oxide from its homeostatic vascular function.

Unstable hemoglobin variants

There is a variety of different mechanisms underlying hemoglobin stability resulting from amino acid substitutions in different parts of the molecule. The first is typified by amino acid substitutions in the vicinity of the heme pocket, all of which lead to a decrease in stability of the binding of heme to globin. A second group of unstable variants results from amino acids that simply disrupt the secondary structure of the globin chains. About 75% of globin is in the form of α helix, in which proline cannot participate except as part of one of the initial three residues. At least 11 unstable hemoglobin variants have been described that result from the substitution of proline for leucine, five that are caused by the substitution of alanine by proline, and three in which proline is substituted for histidine. Another group of variants that causes disruption of the normal configuration of the hemoglobin molecule involves internal substitutions that somehow interfere with its stabilization by hydrophobic interactions. Finally, there are two groups of unstable hemoglobins that result from gross structural abnormalities of the globin subunits; many are due to deletions involving regions at or near interhelical corners. A few of the elongated globin chain variants are also unstable.

Abnormal oxygen transport

There is a family of hemoglobin variants associated with high oxygen affinity and hereditary polycythemia. Most result from amino acid substitutions that affect the equilibrium between the R and T states (*see section Structure and function*). Thus, many of them result from amino acid substitutions at the α_1–β_2 interface, the C-terminal end of the β chain, and at the 2,3-DPG binding sites.

Congenital cyanosis due to hemoglobin variants

There is a family of structural hemoglobin variants that is designated Hb M, to indicate congenital methemoglobinemia, and is further defined by their place of discovery. The

iron atom of heme is normally linked to the imidazole group of the proximal histidine residue of the α and β chains. There is another histidine residue on the opposite side, near the sixth coordination position of the heme iron; this, the so-called distal histidine residue, is the normal site of binding of oxygen. Several M hemoglobins result from the substitution of a tyrosine for either the proximal or distal histidine residue in the α or β chain.

Postscript

In this short account of the molecular pathology of hemoglobin we have considered how mutations at or close to the α or β globin genes result in a diverse family of clinical disorders due to the defective synthesis of hemoglobin or its abnormal structure. Work in this field over the last 30 years has given us a fairly good idea of the repertoire of different mutations that underlie single-gene disorders and how these are expressed as discrete clinical phenotypes. Perhaps more importantly, however, the globin field has taught us how the interaction of a limited number of genes can produce a remarkably diverse series of clinical pictures, and something of the basis for how monogenic diseases due to the same mutation may vary widely in their clinical expression.

Further reading

Bank A. (2006) Regulation of human fetal hemoglobin: new players, new complexities. *Blood*, **107**, 435–443.

Gibbons RJ, Wada T. (2004) ATRX and X-linked (alpha)-thalassemia mental retardation syndrome. In: Epstein CJ, Erickson RP, Wynshaw-Boris A (eds). *Inborn Errors of Development*. Oxford: Oxford University Press, pp. 747–757.

Higgs DR. (2004) Ham-Wasserman lecture: gene regulation in hematopoiesis: new lessons from thalassemia. *Hematology. American Society of Hematology Education Program*, 1–13.

Steinberg MH, Forget BG, Higgs DR, Weatherall DJ. (2008) *Disorders of Hemoglobin*, 2nd edn. Cambridge: Cambridge University Press.

Weatherall DJ. (2001) Phenotype–genotype relations in monogenic disease: lessons from the thalassaemias. *Nature Reviews. Genetics*, **2**, 245–255.

Weatherall DJ. (2004) Thalassaemia: the long road from bedside to genome. *Nature Reviews. Genetics*, **5**, 625–631.

Weatherall DJ, Clegg JB. (2001) *The Thalassaemia Syndromes*. Oxford: Blackwell.