PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Genetic Disorders of Human Haemoglobin as Models for Analysing Gene Regulation

D. J. Weatherall, D. H. Higgs, W. G. Wood and J. B. Clegg

Phil. Trans. R. Soc. Lond. B 1984 **307**, 247-259
doi: 10.1098/rstb.1984.0125

Phil. Trans. R. Soc. Lond. B **307**, 247–259 (1984) [247]
Printed in Great Britain

Genetic disorders of human haemoglobin as models for analysing gene regulation

By D. J. Weatherall, F.R.S., D. H. Higgs, W. G. Wood and J. B. Clegg
M.R.C. Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, U.K.

The genetic and acquired disorders of human haemoglobin provide a diverse group of naturally occurring models for analysing the regulation of protein synthesis. They include structural haemoglobin variants, thalassaemias, which are conditions in which there is a reduced rate of globin chain production, and hereditary persistence of foetal haemoglobin (HPFH) in which there is an inherited abnormality in the switch from foetal to adult haemoglobin synthesis. The thalassaemias result from a diverse series of *cis* acting lesions of the globin genes which include deletions, insertions, frame shift mutations, and point mutations involving transcription, messenger RNA processing, initiation, termination, poly A addition and globin chain stability. Many forms of HPFH are due to deletions of the β-like gene cluster; it has been suggested that they may involve regions of the cluster which are involved in the regulation of the foetal to adult globin chain switch. So far, however, no regions of this type have been identified with certainty. The varieties of HPFH not associated with major gene deletions, or those caused by genetic determinants that are not linked to the globin gene clusters, and some of the acquired forms of α thalassaemia associated with mental retardation or leukaemia, may be more useful models for studying the regulation of the globin genes, particularly their developmental control.

From work done over the last few years it is apparent that the human haemoglobin disorders have a remarkably diverse molecular pathology and that some of them provide useful models for analysing gene regulation. In this brief review we shall summarize recent developments in this field and highlight those conditions that seem to be of particular interest as models for understanding the control of gene expression. For more extensive discussion of this topic, and for reference to much of the original work on which this summary is based, see Maniatis *et al.* (1980), Higgs & Weatherall (1983), Wood & Weatherall (1983), Orkin *et al.* (1983), and Collins & Weissman (1984).

The globin gene clusters

The human haemoglobins and the arrangement of their genes are summarized in figure 1. The β-like globin genes form a linked cluster on chromosome 11 which is spread over approximately 60 kilobases; they are arranged in the order 5′–ε–$^{G}\gamma$–$^{A}\gamma$–ψβ–δ–β–3′. The α-like globin genes form a smaller cluster on chromosome 16, in the order 5′–ζ–ψζ–ψα–α2–α1–3′. The ψβ, ψζ and ψα genes are pseudogenes. The position of the introns, which are found in all the globin genes, is shown in figure 1. The 5′-flanking regions of each of the genes contain two regions of homology. One, the ATA box, is 20–30 base pairs (b.p.) upstream from the RNA initiation (CAP) site. The other, the CCAAT box, is 70–90 b.p. upstream from this site.

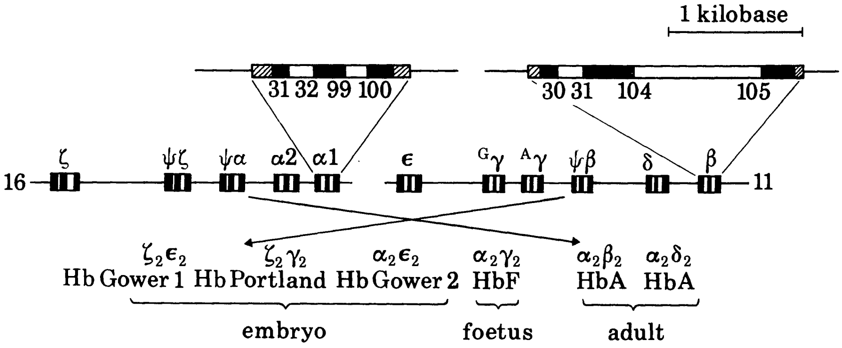

FIGURE 1. The genetic control of human haemoglobin.

Each α gene is located within a region of homology approximately 4 kilobases long, interrupted by two small non-homologous regions. It is thought that the homologous regions have resulted from a duplication event and that the non-homologous segments have arisen subsequently by insertion of DNA into the non-coding region round one of the two genes. The exons of the two α globin genes have an identical sequence. The first intron in each gene is identical but the second intron of α1 is 9 bases longer and differs by 3 bases from the same region in the α2 gene. Despite the high degree of homology between the two genes, the sequences diverge in the 3′ untranslated regions, 13 bases beyond the TAA stop codon. These sequence differences provide an opportunity to assess the relative output of the two genes by DNA–mRNA hybridization; the production of α2 mRNA exceeds that of α1 mRNA by a factor of 1.5–3.0. The ψα gene contains a deletion which brings the CCAAT and ATA boxes unusually close together. In transcription systems *in vitro* it is functional at about 10% of α. A mutation in the poly A addition site may explain the absence of ψα transcripts in reticulocytes.

The two ζ genes are also highly homologous. However, the introns are much larger than those of the α globin genes and, in contrast to the latter, IVS 1 is larger than IVS 2. In each ζ gene, IVS 1 contains several copies of a simple repeated 14 b.p. sequence (ACAGTGGGGA-GGGGG) which is similar to a sequence located between the two ζ genes (see below) and near the human insulin gene. There are three base changes in the coding sequence of the first exon of ψζ, one of which gives rise to a premature stop codon.

The regions separating and surrounding the α-like structural genes have also been analysed in detail. Alu repeat sequences are found around each of the α globin regions; recent studies suggest that they have been dispersed into these regions at two different times in evolution. Two other regions of particular interest have been found flanking the α gene complex, one downstream from the α1 gene and the other between the ζ genes. Their lengths are highly polymorphic. The basic structure of the 3′ hypervariable region consists of a directly repeating 36 nucleotide (5′–GGGGCACAGGTTGTGAG[A or G]GTGCCCGGGACGGCTTGT–3′) segment of DNA. This consists of two 14 b.p. regions separated by a spacer region. The 14 b.p. domains show considerable homology to the repeat sequences in the IVS 1 of the ψζ gene and the hypervariable region near the human insulin gene. Rearrangements involving these repeats are responsible for the length polymorphisms. In addition to the hypervariable regions, several other single base restriction fragment length polymorphisms (RFLPs) have been found in the α gene cluster.

The β globin gene cluster occupies a region of approximately 60 kilobases on the short arm of chromosome 11. Each of its constituent genes, their flanking regions, and large stretches of the regions between them have been sequenced. Like the α1 and α2 gene pairs, the $^G\gamma$ and $^A\gamma$ genes appear to be virtually identical, suggesting a mechanism for gene matching during evolution. The $^G\gamma$ and $^A\gamma$ genes on one chromosome are identical in the region 5′ to the centre of the large intron, yet show greater divergence 3′ to that position. Examination of the boundary between the conserved and divergent regions reveals a block of simple sequence. It has been suggested that this is a 'hotspot' for the initiation of recombination events which lead to unidirectional gene conversion.

Several classes of repetitive sequences have been identified in the γδβ globin gene cluster. First, there are Alu repeat sequences. There are single Alu sequences upstream from the γ globin genes and from the β gene, and inverted pairs of Alu sequences upstream from the ε and δ genes and downstream from the β globin gene. These three inverted pairs are all arranged tail to tail with about 800 b.p. of non-repetitive DNA between them. The second major class of repeated sequences belong to the Kpn family. One copy lies downstream from the β globin gene; another lies between the ε and γ genes. The latter, which is over 6 kilobases in length, has been sequenced, and at the end near the γ globin gene has structural homology to a retrovirus long-terminal repeat.

Apart from a region of short repeats upstream from the β locus, no hypervariable regions have been found in the γδβ globin gene cluster although there are single point RFLPs scattered throughout its length.

Genetic disorders of haemoglobin

The genetic disorders of haemoglobin are divided into the structural haemoglobin variants and the thalassaemias. The thalassaemias are all characterized by a reduced rate of synthesis of one or more of the globin chains. They are classified into the α, β, δβ, and εγδβ thalassaemias respectively. In addition, there is a group of mutations of the globin genes, related to the thalassaemias, which interfere with the switching of foetal to adult haemoglobin production; these conditions are known collectively as 'hereditary persistence of foetal haemoglobin' (HPFH).

In general, studies of the structural haemoglobin variants have told us less than was originally hoped about the regulation of protein synthesis. Shortly after they were first discovered it was noted that many of these variants are found in the red cells of heterozygotes at levels that differed from that of haemoglobin A. This led to several structure–rate hypotheses in which it was suggested that mutations that cause structural haemoglobin variants might have reduced rates of production as compared with haemoglobin A because they utilize transfer RNAs which are in relatively short supply. More recently, it has become apparent that this is probably not the case and that most variant haemoglobins are synthesized at approximately the same rate as normal haemoglobin; their final level in the red cell is determined by fine tuning at the level of sub-unit association. A few structural haemoglobin variants are produced at unusually low rates and are associated with the clinical phenotype of thalassaemia. We shall consider the molecular mechanisms for this phenomenon in the next section.

It is the elucidation of the molecular basis for the thalassaemias that has told us most about the mechanisms involved in the regulation of human gene expression. In addition, studies of the sites and patterns or foetal haemoglobin synthesis, and of the molecular basis of hereditary persistence of foetal haemoglobin, have provided some information about the tissue specificity and regulatory mechanisms involved in the developmental switch from foetal to adult haemoglobin production.

The molecular pathology of the thalassaemias

The α thalassaemias (see Higgs & Weatherall 1983)

There are two important clinical forms of α thalassaemia, the Hb Bart's hydrops foetalis syndrome and Hb H disease. The former is characterized by severe intrauterine hypoxia with stillbirth, or death just after birth. Affected infants are profoundly anaemic and their red cells show only Hbs Bart's (γ_4) and Portland ($\zeta_2\gamma_2$). Haemoglobin H disease is a milder condition compatible with survival into adult life. There is a moderate degree of anaemia and the reduced rate of α chain production leads to the accumulation of excess β chains which form β_4 tetramers, or Hb H. These conditions result from the interaction of two types of α thalassaemia determinants, α^+ and α^0 thalassaemia. The Hb Bart's hydrops syndrome results from the homozygous inheritance of α^0 thalassaemia, while Hb H disease usually results from the compound heterozygous inheritance of α^0 and α^+ thalassaemia.

α^0 thalassaemia is characterized by a complete absence of α chain production. On the other hand, a chromosome with an α^+ thalassaemia determinant directs some α chain synthesis but at a reduced rate. The simplest way in which these two conditions might arise is shown in figure 2. In this model, α^0 thalassaemia results from the deletion of both linked α globin genes, whereas α^+ thalassaemia results from the deletion of only one of the pair of α genes. In fact all the α^0 thalassaemias are due to deletions which remove both α globin genes. However, the α^+ thalassaemias are more complicated. In some cases the genes are deleted, whereas in others they are intact but have mutations which partly or completely inactivate them. Hence, the α^+ thalassaemias are divided into deletion and non-deletion types.

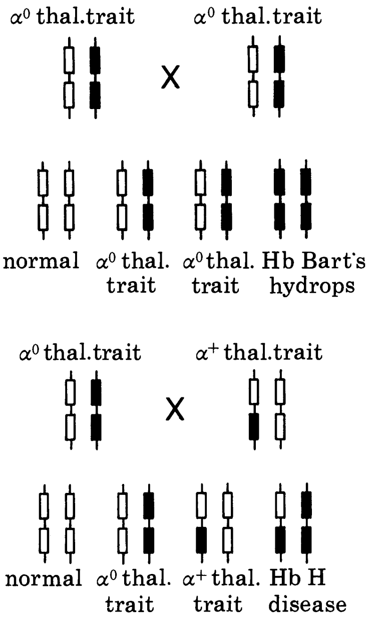

Figure 2. A model for the patterns of inheritance of α^0 and α^+ thalassaemia. The light boxes indicate normal α genes while the dark boxes represent either deleted or otherwise inactivated α genes.

The α^0 thalassaemias are caused by different length deletions of the α globin gene cluster. They start downstream from or within the $\alpha1$ globin genes (figure 3) and extend upstream through the α globin gene cluster removing the $\alpha2$ gene, and, in some cases, the $\psi\zeta$ gene; each of them leaves the functional ζ gene intact. This is why infants with the haemoglobin Bart's hydrops syndrome can produce Hb Portland ($\zeta_2\gamma_2$), and hence survive to term.

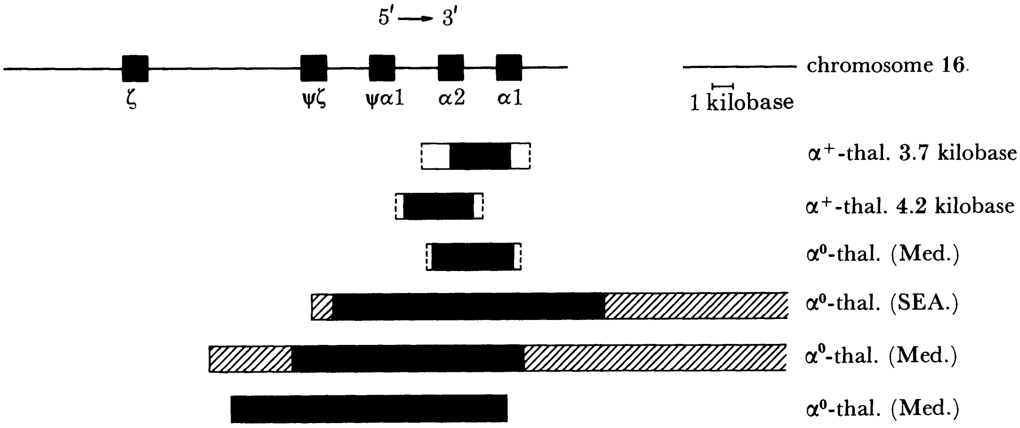

FIGURE 3. The deletion forms of α^+ and α^0 thalassaemia. Med: Mediterranean; SEA; Southeast Asia. The dark regions indicate deleted areas of the gene cluster and the hatched regions indicate uncertainty about the precise length of a deletion.

In the deletion forms of α^+ thalassaemia, different-sized deletions remove one α globin gene and leave one functional gene ($-\alpha$) (figure 3). These lesions have arisen by unequal crossing over between homologous pairs of chromosomes 16, leaving one α globin gene on one of the pair and three on the other. One deletion, which removes 3.7 kilobases, involves both genes and leaves a single composite α gene; the other, a 4.2 kilobase deletion, removes the $\alpha2$ gene. Individuals with three α globin genes on one chromosome and two on the other, i.e. five α globin genes in all, have been found in many different populations. The three α globin genes may represent the anti-3.7 or anti-4.2 crossover chromosomes. Homozygotes, with six α globin genes, have also been encountered.

Several different types of 'non-deletion' α thalassaemia have been defined. In one, 5 bases (TGAGG) following the G of the invariant GT within the donor splice site are lost; this completely inactivates the $\alpha2$ gene. In another, there is a single base mutation (GTG → CCG in the $\alpha2$ gene) which produces a highly unstable α chain variant, haemoglobin Quong Sze, which is rapidly destroyed and hence results in an α thalassaemia phenotype. Recently, a novel type of non-deletion α thalassaemia, which occurs commonly in Middle Eastern populations, has been characterized. In this case the coding sequence of the $\alpha2$ gene is completely normal apart from a mutation in the highly conserved AATAAA sequence in the non-coding region, which has changed to AATAAG. When this gene was analysed in a transient expression system *in vitro* it was found that, while some normal sized α globin mRNA was produced, a considerable proportion remained unprocessed as a longer precursor. It is believed that the AATAAA sequence is the recognition site for polyadenylation. Presumably this mutation reduces the efficiency of this process and hence a proportion of the α globin mRNA precursor is not cleaved and polyadenylated. Hence the output of the $\alpha2$ gene is markedly reduced. However, since this mutation was found on a chromosome in which the α1 gene had a frameshift mutation, and since there were some normal α globin gene chains produced in individuals homozygous for these lesions, it is clear that the polyA addition signal mutation is associated with the production of some normal gene product.

Another group of non-deletion α thalassaemias result from single base changes in the α globin chain termination codon. These mutations allow an amino acid to be inserted at what is normally the position of the stop codon, and then sequences at the 3′ end of the α globin messenger RNA which are not normally utilized are translated. This results in the production of α globin chain variants with 31 additional amino acid residues at their C terminal ends; the prototype is Hb Constant Spring in which the change in the termination codon, UAA to CAA, is reflected by the glutamine residue which is the first amino acid in the elongated part of the α chain. Recent studies indicate that the mutation occurs in the α2 locus. Since no α2 globin mRNA is found in the peripheral blood cells of individuals with this disorder, and since the α chains of Hb Constant Spring are synthesized in early red cell precursors but not in peripheral blood reticulocytes, it appears that the translation of sequences at the 3′ end of the α globin messenger RNA which are not normally utilized render it unstable. Hence, haemoglobin Constant Spring is produced in very low quantities and is associated with the phenotype of α^+ thalassaemia. These findings suggest that there must be sequences in the 3′ untranslated region of α globin messenger RNA which are involved in its stabilization; presumably the disruption caused by 'read through' of these sequences interferes with this process.

Analysis of the α thalassaemias has also provided useful information about translational and post-translational regulation of protein synthesis. Although messenger RNA analyses indicate that the output of the α2 gene exceeds that of the α1 gene, it appears that fine tuning at the translational level leads to a similar output of α globin chains from the two loci. As judged by both globin chain synthesis and globin messenger RNA analyses, deletions of one, two and three α globin genes are characterized by a deficiency of α globin chains, the magnitude of which is compatible with each α gene contributing more-or-less equally to the overall amount of α globin.

The finding of triplicated α globin gene arrangements, and that in most cases the three α globin genes are functional, provides further insight into the mechanisms that determine the amount of mutant haemoglobins in the red cells of heterozygotes. For example, individuals heterozygous for β chain haemoglobin variants such as Hb S usually have lower levels of the variant than of Hb A. It has been suggested that, although β^A and β^S chains are synthesized at equal rates, because β^S chains combine with α chains less efficiently than with β^A chains, some β^S chains are rapidly destroyed after synthesis; hence the final amount of Hb A exceeds that of Hb S. Recent observations on Hb S heterozygotes who have inherited either one, two or three α globin genes together are in keeping with this model. There is a strong correlation between the number of α globin genes and the level of Hb S; the greater the number of α genes the higher the level of Hb S. If there is competition for α chains between β^A and β^S chains, the larger the amount of α chains the greater likelihood there is that they will combine with β^S chains.

Finally, these structural studies of the α globin gene cluster raise certain important questions about the arrangement of the α globin genes. Why do all species studied to date have duplicated and structurally identical α globin genes? Although the answer to this important question is not known, analysis of the α thalassaemias may provide a clue to the mechanism whereby the α globin genes have been matched during evolution.

The deletion forms of α^+ thalassaemia are widespread and reach very high frequencies in some populations. Interestingly, the triplicated α gene arrangement has also been found in every population looked at so far, though never at a frequency of more than 2%. Triplicated and single ζ gene arrangements also occur in several populations. These observations suggest that unequal crossing over at the α or ζ gene complexes may be a common event. As mentioned earlier, it has been suggested that α globin gene sequence matching could occur by expansion and contraction of gene number by such a process. Thus, the deletion forms of α^+ thalassaemia may have arisen as part of an evolutionary mechanism for α globin gene matching, selection for the single α gene chromosomes being much greater than for the triplicated arrangement.

Whatever the mechanism for the production of the different forms of α thalassaemia, it is clear that the α globin genome is by no means static. Single or triplicated α and ζ gene arrangements are common, there is a number of length polymorphisms in the two hypervariable regions, and there is a diversity in the size of the introns in the ζ globin genes. All these variations, together with many different deletion or non-deletion α thalassaemia mutations, combine to produce a highly variable α globin genome which, except in cases of homozygosity for major gene deletions, retains sufficient functional adaptability for adequate α chain production. So far, no definite function has been ascribed to any of the intergenic regions of the cluster, and the reason for the high degree of conservation of the repeat sequences in the ψζ IVS, and its relationship to the 5′ hypervariable regions, remains unexplained.

The acquired forms of α thalassaemia (*Weatherall et al.* 1982)

The forms of α thalassaemia which we have considered so far are all due to simple *cis*-acting mutations that involve the α globin gene cluster. However, there are some less well defined lesions that involve this cluster which, in the long term, may be of more general interest in terms of gene regulation.

A series of children of North European background have been described with the unusual combination of mental retardation and haemoglobin H disease. The findings in the parents of these children are quite different from those described earlier for the common genetic forms of this condition. In each family, one parent is completely normal while the other has a form of α^+ thalassaemia. It has been suggested, therefore, that these children have inherited an α^+ thalassaemia gene from one parent, and a mutation, probably from the other parental germ cell line, which completely inactivates both α globin genes on chromosome 16. The lesions on chromosome 16 are heterogeneous; in some cases there is a long deletion which removes the entire α-like gene complex, while in others the complex appears to be normal by restriction enzyme analysis and yet both α genes are inactive. It has been suggested that the defect on chromosone 16 may be related to the associated mental retardation and that these children only presented with a haematological disorder because they happened to inherit an α^+ thalassaemia trait. This hypothesis has been strengthened recently by the discovery of a mentally retarded child with the phenotype of α^0 thalassaemia trait, both of whose parents were completely normal. This child appears to have acquired a defect which has shut down the activity of both α globin genes; there are no structural abnormalities of the α gene complex. It will be of great interest to determine the molecular basis for these acquired forms of α^0 thalassaemia in which the α gene complex is normal by gene mapping.

There is another acquired form of α thalassaemia which is associated with leukaemic transformation of the bone marrow in elderly patients. There is a gradual emergence of a cell line in which there is complete suppression of α globin chain synthesis, yet both pairs of α globin genes are intact. It seems likely that the molecular basis for this condition involves a gene, or genes, which have a role in regulating the entire α globin gene cluster.

The β thalassaemias (*Orkin et al.* 1982; 1983)

The β thalassaemias are characterized by a reduced rate of β chain production (β^+ thalassaemia), or an absence of β chain production (β^0 thalassaemia).

Overall, the molecular pathology of the β thalassaemias is different from that of the α thalassaemias. Deletions are the exception rather than the rule; only one common form of β thalassaemia due to a major deletion has been found. This lesion, which occurs in some Indian populations, results from the loss of 619 bases at the 3′ end of the β globin gene; the deletion starts in the large intron and extends downstream beyond the coding region of the β gene. The remainder of the β thalassaemia mutations which have been defined so far are single base changes, or small deletions or insertions in the β globin genes or their flanking regions. The main classes of β thalassaemia mutations are summarized in table 1.

Table 1. The main groups of β thalassaemia

(1) chain termination (nonsense) mutations
(2) frameshift mutations
(3) deletions
(4) mutations causing defective RNA processing
  splice junction
  consensus region
  internal IVS
  cryptic splice site in exons
  poly A addition site
(5) mutations causing inefficient transcription

Two main types of mutations lead to the production of non-functional β globin mRNA. Single base changes in codons 17 or 39 produce premature stop codons. In addition, there are at least six frameshift mutations which result from the loss or insertion of 1, 2 or 4 bases in codons 6, 8, 8–9, 16, 41–42, 44, and 71–72. The nonsense and frameshift mutations all produce the clinical phenotype of β^0 thalassaemia.

The RNA processing mutations fall into four classes (table 1). First, there are the splice junction mutations which involve the invariant GT donor sites or AG receptor sites at the IVS 1 or IVS 2 intron–exon junctions. There is one variant that results from a deletion of 25 bases at the acceptor site at the 3′ end of IVS 1. Each of these mutations completely abolishes normal splicing. The second group comprises two mutations that involve single base changes in the IVS 1 consensus sequences although not within the invariant nucleotides. Both these mutations lead to decreased splicing at the donor site and low-level alternative RNA processing into donor-like sequences in both exon 1 and IVS 1 (figure 4). The third class involves the generation of new splicing signals within an IVS. Four mutations of this type are known. For example, a single base substitution at IVS 1, position 110, produces a new acceptor site; both normal and abnormally spliced messenger RNA are produced and the resulting phenotype is a severe form of β^+ thalassaemia. The fourth class of processing mutants consists of substitutions within coding regions which activate cryptic splice sites. For example, the region of exon 1 which contains codons 24–27, and which is altered in the production of haemoglobin E, is similar to a donor splice site. It is not normally used in RNA processing although it appears to be utilized in processing abnormal RNA species produced from three mutant genes of the IVS 1 donor variety (figure 4). The mutation which produces the amino acid change in the β chain of Hb E alters this cryptic donor site and renders its functional at a low level during mRNA processing. It appears that competition of this exon 1 donor site with the authentic IVS 1 donor site results in delayed as well as alternative β globin mRNA processing. This may be the basis for the thalassaemia-like phenotype associated with Hb E; a similar mechanism may also explain the association of Hb Knossos with a thalassaemic phenotype.

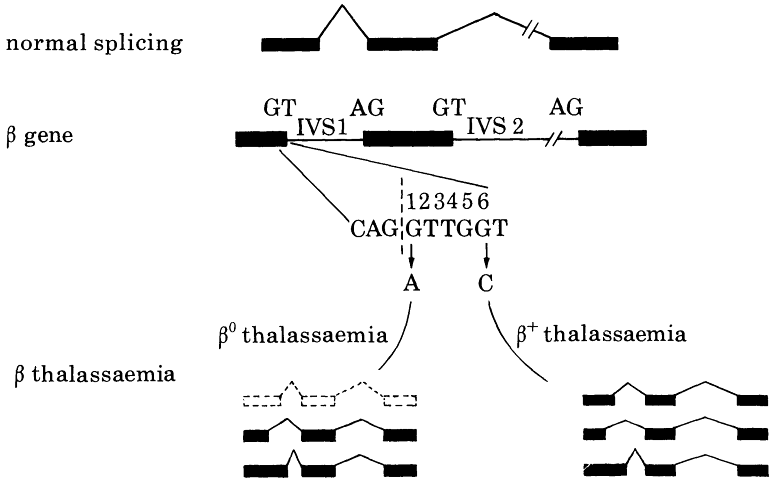

Figure 4. β thalassaemia mutants with substitutions in the donor splice consensus sequence of IVS 1.

Another group of β thalassaemia mutations occurs in the 5′ flanking region of the β globin gene. It leads to inefficient transcription of the β globin gene. The −87 C→G mutation (close to the CCAAT box) occurs widely throughout the Mediterranean region. It, and a recently discovered −88 mutation, is associated with a particularly mild form of β$^+$ thalassaemia. Similarly, two different mutations at −28 and one at −29, in the region of the TATA box, also cause mild forms of β$^+$ thalassaemia.

Finally, a poly A addition signal mutation, AATAAA→AACAAA, causes a mild form of β$^+$ thalassaemia. Presumably the mechanism of reduced β globin chain production is similar to that described earlier for the poly A addition signal mutation which causes a form of α$^+$ thalassaemia.

Thus the β thalassaemias provide a useful series of naturally occurring models for analysing different aspects of the regulation of protein synthesis. The 5′ non-flanking region mutations confirm that the various 'regulatory boxes' are involved in control of transcription of the globin genes; as more of them are defined it should be possible to characterize the relative importance of these different regions in determining rates of transcription. The various splicing mutations provide important information about the specificity of the consensus regions in the donor splice sites and about the existence of cryptic splice sites in exons. As more of these lesions are sequenced it will be interesting to determine how much fine tuning of RNA processing is mediated through subtle sequence differences in these regions. As with the α thalassaemias, all the mutations which have been found so far in β thalassaemia have been *cis*-acting, either within the structural genes or in their flanking regions.

δβ *thalassaemia and hereditary persistence of foetal haemoglobin*

The genes of the γδβ globin gene cluster are arranged in the order 5′→3′, in which they are expressed during foetal development. It has long been suspected that this arrangement may be of functional significance and that an analysis of lesions in this gene cluster associated with persistent γ chain synthesis in adult life might provide some evidence about the mechanisms that regulate these loci during normal development.

Persistent γ chain synthesis occurs in homozygotes, but not heterozygotes, in the β thalassaemias and sickling disorders. However, it seems unlikely that the primary gene mutations that cause these conditions are the direct cause of the apparent increase in γ chain production. Indeed, much of it can be explained by cell selection. As a result, these conditions are not helpful for analysing the regulation of γ gene expression. However, the δβ thalassaemias and HPFH are characterized by persistent high levels of γ chain production in adult life in heterozygotes which cannot be accounted for by cell selection. Hence, the lesions that cause these conditions may involve regions of the γδβ gene cluster involved in the regulation of γ chain production.

The δβ thalassaemias and some forms of HPFH result from long deletions of the γδβ globin gene cluster. All these conditions are characterized by the absence of δ and β globin chain synthesis in homozygotes, who have only Hb F in their red cells. In some cases the deletions also involve the $^{A}\gamma$ genes and the haemoglobin F contains only $^{G}\gamma$ chains. Hence it has been customary to call these conditions $^{G}\gamma$ or $^{G}\gamma^{A}\gamma$ δβ thalassaemia or HPFH. As indicated in figure 5, it is more logical to designate them by the globin chains which are not produced.

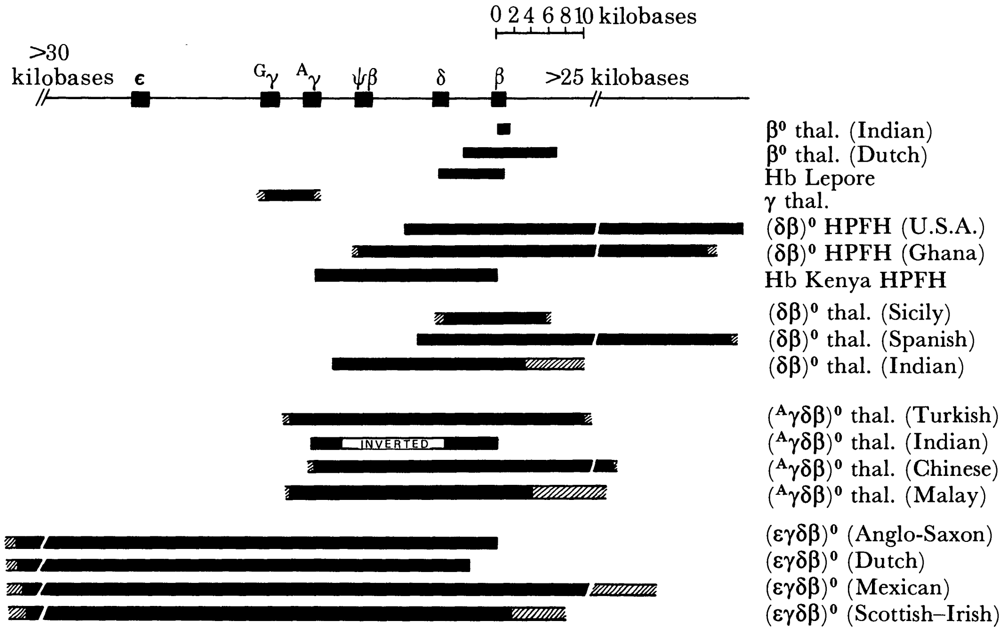

FIGURE 5. Deletions of the β-like globin gene cluster with their associated phenotypes. HPFH: hereditary persistence of foetal haemoglobin.

Thus, $^G\gamma^A\gamma$ δβ thalassaemia should be called $(\delta\beta)^0$ thalassaemia; $^G\gamma$ δβ thalassaemia $(^A\gamma\ \delta\beta)^0$ thalassaemia, and so on.

The δβ thalassaemias and deletion forms of HPFH are all associated with high levels of γ chain synthesis in adult life. This does not entirely compensate for the absence of β and δ chain production however, and these disorders all have a mild thalassaemic phenotype. In the case of HPFH, compensation is almost complete, although even in this condition homozygotes have very mild thalassaemic phenotypes. In δβ thalassaemia there is even less effective compensation for defective β and δ globin chain production, and homozygotes have a moderately severe thalassaemic disorder. For this reason, there has been considerable interest in comparing the extent of the deletions that produce the δβ thalassaemic and HPFH phenotypes.

Some of the different deletions that produce δβ thalassaemia and HPFH are summarized in figure 5. A particularly interesting comparison is the 5′ extent of the deletions which cause either $(\delta\beta)^0$ thalassaemia or $(\delta\beta)^0$ HPFH in Negro populations. These deletions end within 1 kilobase of each other in the Alu repeat region 5′ to the δ globin gene. The deletion that causes HPFH ends in the middle of the upstream Alu repeat while that causing δβ thalassaemia ends about 1 kilobase downstream from the latter in the other Alu repeat. Is this Alu repeat region involved in the regulation of γ and β chain synthesis during development? While this may be the case it should be remembered that both these deletions cause a considerably elevated level of γ chain production in adult life; the difference between them is only a matter of degree. Furthermore, there is no phenotypic difference between the different types of $(\delta\beta)^0$ thalassaemia, yet one of them is caused by a deletion which ends in the δ globin gene. Recently, the DNA sequence of this region has been determined but it seems to contain no particularly unusual features other than the Alu elements themselves. In addition, the sequence between the two Alu sequences have been analysed in DNA from several different individuals. Although an interesting dimorphism has been observed, in which there are two alleles which differ by 16 point mutations and two deletions, these alleles do not correlate with the amount of Hb F synthesized and it seems unlikely that they have a role in haemoglobin switching.

It has been suggested that there are two chromatin domains in the β gene cluster, one surrounding the foetal genes with distinct 5′ and 3′ borders, the other flanking the δ and β genes with similar distinct borders. The idea is that foetal to adult haemoglobin switching occurs when the foetal domain closes and the adult domain opens. It follows that mutations that alter the 5′ border of the δβ domain still allow some switching to occur and result in a thalassaemia phenotype; mutations that affect both 5′ and 3′ δβ domains result in persistent over-expression of the foetal domain and hence in the HPFH phenotype. Another suggestion is that what is important for the regulation of gene expression in these deletion mutations is not the region of DNA deleted but the sequences brought into apposition to the β globin gene complex by the deletion. Perhaps in some forms of HPFH, sequences are brought in from 3′ to the β globin complex that act as *cis* enhancers, thus allowing expression of the foetal genes in adult life. However, as shown in figure 5, the 3′ end of all these deletions are different; do they all contain enhancer sequences? Furthermore, the form of $(^A\gamma\ \delta\beta)^0$ thalassaemia caused by an inversion of the sequence between the $^A\gamma$ and δ genes has normal sequences 3′ to the β gene, yet the $^G\gamma$ genes are active. Thus, at the present time it is impossible to distinguish between these models, or even to be sure that they are mutually exclusive; none of them is compatible with all the phenotypic findings.

Despite the fact that no regulatory regions have been definitely defined by these deletions, they provide some useful information, particularly when compared with those that involve the α-like globin gene cluster on chromosome 16 (figure 3). It appears from an analysis of the deletions of the β-like gene cluster that downstream lesions are usually associated with persistence of upstream loci which are normally turned off during foetal development. However, in the deletions which involve the α-like or β-like clusters embryonic globin chain synthesis does not persist. Thus it appears that however much these clusters are disrupted by long deletions, the embryonic globin genes remain inactive. In the context of the domain theory outlined above, it is possible that the embryonic globin genes in both clusters belong to separate domains that are not modified by these deletions; the effect of deletions further downstream may, as suggested above, depend on their precise relationship to the foetal (γ globin gene) and adult (β and δ globin gene) domains.

Perhaps further studies of these extensive deletions, because of the disruption of the cluster that they must cause, may be of limited value for defining mechanisms involved in the regulation of globin gene switching. However, there are at least four well defined types of non-deletion HPFH, the Greek, Chinese, British and Negro $^{G}\gamma\ \beta^{+}$ varieties. In the British form, in which both heterozygotes and homozygotes have been identified, there is persistent $^{A}\gamma$ chain synthesis with β globin chain production *cis* to the genetic determinant. The Greek and Chinese forms seem to be similar although there are higher levels of $^{A}\gamma$ chains in heterozygotes. Detailed gene mapping has shown no abnormality in these conditions. In $^{G}\gamma\ \beta^{+}$ HPFH, heterozygotes produce approximately 20% Hb F of the $^{G}\gamma$ type and there is β chain production in *cis*. This condition has been analysed in three different laboratories. Extensive sequence analysis of the $^{A}\gamma$ gene and of the Alu repeat region 5′ to the δ gene has shown no abnormality. However, a single point mutation has been identified 202 base pairs 5′ to the CAP site of the $^{G}\gamma$ gene. This occurs in a GC rich palindrome and creates a sequence that resembles the herpes thymidine kinase and SV40-21 base pair repeat promotor elements. It has been suggested that the −202 mutation occurs in a similar element which regulates the γ genes and hence leads to their increased transcription in adult life. This mutation has been found in at least three individuals with $^{G}\gamma\ \beta^{+}$ thalassaemia but the possibility remains that it is a simple base polymorphism.

Finally, there is a group of non-deletion HPFH-like conditions associated with low levels of Hb F which is heterogeneously distributed among the red cells. Although some of them are caused by determinants which, as judged by RFLP linkage analysis, map within the γβ globin gene cluster, several families have now been reported in which this is not the case. This is the first evidence for the existence of genes which influence globin chain synthesis but which are not linked to the β-like globin gene cluster.

Studies of the deletion mutants suggest that there must be some form of *cis* regulation in the γδβ gene cluster; available evidence suggests that γ chain production occurs only in *cis* to these mutations. As yet, there is no evidence as to whether this is the case for the non-deletion forms of HPFH, and so far no examples of *cis–trans* acting mutations have been discovered. However, recent gene or chromosome transfer experiments suggest that *trans* regulators may exist, and that they may be stage specific. For example, it has been found that transfection of mouse erythroleukaemia (MEL) cells with human chromosome 11, or with cosmids containing human β, γ and ε genes results in a considerable increase in β gene expression after induction of haemoglobin synthesis, whereas there is no increase in the expression of the γ or ε genes. Similarly, when intact human chromosomes 16 are transferred into MEL cells there is expression of the α globin genes but no expression of the embryonic ζ globin genes. Perhaps further experiments of this kind, together with a detailed analysis of the chromosomal origin of γ chains in some of the forms of non-deletion HPFH whose genetic determinants are unlinked to the γδβ globin gene cluster, together with defining the chromosomal location of these non-linked mutations, offers the most promising approach to defining the way in which the γ and β globin genes are regulated during development.

References

Collins, F. S. & Weissman, S. M. 1984 The molecular genetics of human hemoglobin. In *Progress in nucleic acids research and molecular biology*. (In the press.)

Higgs, D. R. & Weatherall, D. J. 1983 Alpha thalassaemia. In *Current topics in hematology* (ed. S. Piomelli & S. Yachnin), pp. 37–97. New York: Alan R. Liss.

Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. 1980 The molecular genetics of human hemoglobins. *A. Rev. Genet.* **14**, 145–178.

Orkin, S. H., Antonarakis, S. E. & Kazazian, H. H. 1983 Polymorphism and molecular pathology of the human beta-globin gene. In *Progress in hematology*, XIII (ed. E. B. Brown), pp. 49–73. New York: Grune & Stratton.

Orkin, S. H., Kazazian, H. H., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G. & Giardina, J. V. 1982 Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. *Nature, Lond.* **296**, 627–631.

Weatherall, D. J., Higgs, D. R., Clegg, J. B. & Wood, W. G. 1982 Annotation. The significance of haemoglobin H disease in patients with mental retardation or myeloproliferative disease. *Brit. J. Haemat.* **52**, 351–355.

Wood, W. G. & Weatherall, D. J. 1983 Developmental genetics of the human haemoglobins. *Biochem. J.* **215**, 1–10.