Cell, Vol. 29, 7–9, May 1982, Copyright © 1982 by MIT

Thalassemia Revisited

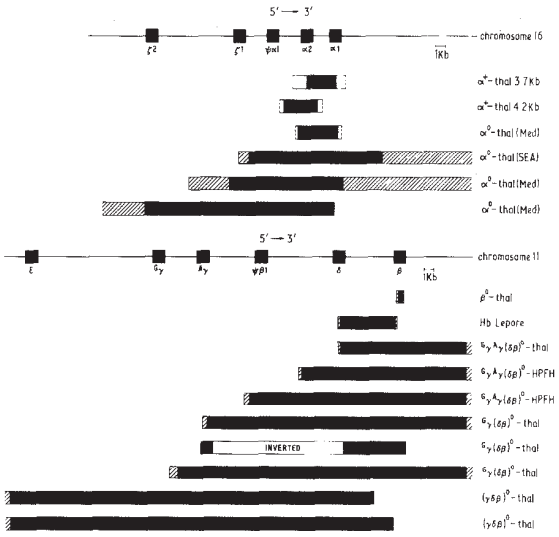
David J. Weatherall and John B. Clegg
MRC Molecular Haematology Unit
Nuffield Department of Clinical Medicine
University of Oxford
John Radcliffe Hospital
Headington, Oxford OX3 9DU, England

Apart from their intrinsic medical importance as the commonest group of monogenic disorders in the world population, the thalassemias provide a variety of naturally occurring models for the study of the regulation of hemoglobin synthesis and its developmental genetics. In the relatively short time since the topic was last reviewed (Weatherall and Clegg, Cell *16*, 467–479, 1979) there has been spectacular progress in the study of human hemoglobin. The fine structure of the globin-gene complex is known, and the molecular bases for several different forms of thalassemia have been determined.

The hallmark of the thalassemias is defective synthesis of the α- or β-like chains of the normal hemoglobin tetramers (hemoglobins A, $\alpha_2\beta_2$; A2, $\alpha_2\delta_2$; and F, $\alpha_2\gamma_2$). The resulting imbalance (that is, of α over β, δ and γ chains, or vice versa) is responsible for the characteristic hematological abnormalities, through the effects of excess free globin chains on red cell maturation and survival. Depending on which globin chain is affected, the disorders are thus classified into α, β, δ, $\delta\beta$ and $\gamma\delta\beta$ thalassemias. Each of these is heterogeneous. For example, α thalassemias can be divided into α^+, in which there is a reduced output of α chains, and α^0 types, with no α-chain synthesis; likewise with the β thalassemias. Within these two broad classes are further subtle phenotypic subdivisions. There is, in addition, a group of closely related conditions called hereditary persistence of fetal hemoglobin (HPFH), in which hemoglobin F production persists into adult life.

The α-globin genes, a pseudo-α gene ($\psi\alpha1$) and their embryonic ζ-gene counterparts (one of which, $\zeta1$, may also be a pseudogene), lie in a cluster on the short arm of chromosome 16 in the order 5′–$\zeta2$–$\zeta1$–$\psi\alpha1$:$\alpha2$–$\alpha1$–3′ (Lauer et al., Cell *20*, 119–130, 1980). There are two highly variable regions in this part of the genome, one 3′ to the α genes and the other between the two ζ genes (Higgs et al., NAR *9*, 4213–4224, 1981). With the possible exception of some cases of hemoglobin H disease associated with mental retardation, all the α^0 thalassemias characterized so far result from gene deletions. In α^0 thalassemia from Southeast Asia a deletion of at least 17.5 kb removes both α genes and the $\psi\alpha$ gene, but leaves both ζ genes intact, while in the common form of α^0 thalassemia found in the Mediterranean the deletion also includes the $\zeta1$ gene (Pressley et al., PNAS *77*, 3586–3589, sults from a small deletion (5.2 kb) that has removed the $\alpha2$ gene and only the 5′ end of the $\alpha1$ gene (Pressley et al., NAR *8*, 4889–4894, 1980); the other involves a deletion that extends at least 25 kb from the middle exon of the $\alpha1$ gene and removes the $\alpha2$ gene, the $\psi\alpha$ gene and both ζ genes (Orkin and Michelson, Nature *286*, 538–540, 1980). There is no indication of how any of these deletions might have arisen. Infants who are homozygous for the common Mediterranean or Southeast Asian forms of α^0 thalassemia produce significant amounts of hemoglobin Portland ($\zeta_2\gamma_2$) in the later stages of gestation. Thus when there are major deletions downstream in the ζ–α globin gene cluster, upstream loci that are normally shut off early during development remain active.

The α^+ thalassemias are also extremely diverse. Both deletion and nondeletion types have been described; the latter appear to have a relatively more severe phenotype. Two common deletion forms of α^+ thalassemia result from the loss of 3.7 and 4.2 kb of the α-gene cluster, leaving one α-globin gene intact and functional (Kan et al., Blood *54*, 1434–1438, 1979; Orkin et al., Cell *17*, 33–42, 1979). The finding of individuals with a triplicated α-gene haplotype corresponding to both these deletions indicates that, like the hemoglobin Lepores, they have arisen by nonhomologous interchromosomal crossing-over (Goosens et al., PNAS *77*, 518–521, 1980; Higgs et al., Nature *284*, 632–635, 1980; Trent et al., Br. J. Haematol. *49*, 149–152, 1981). In Jamaica, about 35% of the population have the $\alpha^{3.7}$ deletion haplotype, whereas only 2% have the $\alpha\alpha\alpha^{anti-3.7}$ haplotype, suggesting that the former has come under much stronger selective pressure. Since both these variants occur in many



Deletions in various thalassemia haplotypes. In the 3.7, 4.2 and 5.2 kb α-gene deletions the vertical dashed lines indicate the limits within

racial groups, these crossovers have probably occurred independently on a number of occasions. Although there are two major regions of homology containing the $\alpha 1$ and $\alpha 2$ genes, it is not yet clear whether recombination has occurred at many places throughout these regions, or whether it is confined to only a few "hotspots."

The nondeletion forms (as defined by restriction mapping of genomic DNA) of α thalassemia are characterized by reduced production of α globin, although both α genes are intact. Quantitation of α mRNA and assessment of the relative proportions of the transcripts of the $\alpha 1$ and $\alpha 2$ loci by Berk-Sharp analysis suggest considerable heterogeneity at the molecular level (Higgs et al., PNAS *78,* 5833–5837, 1981; Liebhaber and Kan, J. Clin. Invest. *68,* 439–466, 1981; Orkin and Goff, Cell *24,* 345–351, 1981). In some instances only one of the pair of α-globin genes is active; sequence analysis of a case of this type has shown a 5 bp deletion in the 5′ IVS1 splice junction that may interfere with normal RNA processing (Orkin et al., PNAS *78,* 5041–5045, 1981). In others both α loci are active, although at a reduced (and variable) level. Some of these conditions may be transcriptional mutations, although others will probably turn out to be nonsense mutations and unstable α variants, as has been found to be the case in some of the β thalassemias. The α-globin chain-termination mutants, such as hemoglobin Constant Spring, have the phenotype of α^+ thalassemia because the α^{CS} mRNA is very unstable. This probably results from nucleolytic degradation following destabilization of the mRNA during readthrough of the normally untranslated 3′ end. These insights into the molecular pathology of α thalassemia emphasize its extraordinary molecular diversity.

Some interesting questions for the population geneticist arise from these observations. What, for example, are the selective factors that have allowed the deletion forms of α^+ thalassemia to reach a prevalence of 30%–40% in parts of West Africa and even as high as 90% in some pockets in India? Why is α^0 thalassemia so rare in these populations when it is very common in Southeast Asia, where there is also a high incidence of the deletion forms of α^+ thalassemia? Why is it that in eastern Saudi Arabia, where well over 50% of the population are affected with both deletion and nondeletion α^+ thalassemia, α^0 thalassemia is unknown? Why in the vast populations of Southeast Asia is there apparently only one α^0-thalassemia haplotype, whereas in the Mediterranean, where α^0 thalassemia is much less common, there are at least three?

A rather different story is emerging for the β thalassemias. With the exception of a variety due to a deletion of the 3′ end of the β-globin gene that is found in some Indians, the β thalassemias seem to from highly unstable β-globin variants. Nonsense mutations, notably those of codons 17 and 39 of the β-globin mRNA, have been found to be the basis for a number of β^0 thalassemias (Chang and Kan, PNAS *76,* 2886–2889, 1979; Orkin and Goff, JBC *256,* 9782–9784, 1981). One example of β^0 thalassemia due to a frameshift mutation has been found (Orkin and Goff, JBC *256,* 9782–9784, 1981). Several cases of β^+ thalassemia are due to a single nucleotide substitution (G to A) in IVS1 of the β gene (Spritz et al., PNAS *78,* 2455–2459, 1981; Westaway and Williamson, NAR *9,* 1777–1788, 1981). This has been cloned into an SV40–pBR328 vector and introduced into HeLa cells, and the RNA produced by the transfected cells has been analyzed by S1 nuclease mapping and cDNA sequencing (Busslinger et al., Cell *27,* 289–298, 1981). By creating an alternative splicing junction, this mutation apparently leads to the production of about 90% of abnormally spliced β-globin mRNA with an early stop codon; however, some mRNA is spliced normally, and hence a β^+ thalassemia phenotype results. In other cases, a G to A substitution in the 5′-end IVS1 or IVS2 splice junctions appears to lead to a more severe processing defect, and hence to the phenotype of β^0 thalassemia (Baird et al., PNAS *78,* 4218–4221, 1981; Orkin et al., Nature, in press). A variety of β^0 thalassemia found in Kurdish Jews is characterized by a highly unstable β mRNA, although the mechanism is not known (Maquat et al., Cell *27,* 543–553, 1981). While such studies have clarified the molecular basis for some types of β^+ and β^0 thalassemia, they have not yet, with one possible exception, unearthed a good candidate for a primary mutation affecting transcription. The exception is the single-base change (C to G) 87 nucleotides preceding the CAP site and just upstream from the CCAAT box in the 5′-flanking region of a β-thalassemia gene (Orkin et al., Nature, in press). Evidence that this is the actual cause of the β thalassemia phenotype has yet to be obtained. The impression gained so far is that mutations that completely block β-gene transcription are extremely rare, in contrast with those defects that affect processing of the primary transcript, or stability or translation of β-globin mRNA.

The β-thalassemic mutations appear to be in linkage disequilibrium with a limited group of restriction enzyme polymorphisms within and flanking the β-gene cluster (Orkin et al., Nature, in press). In the Mediterranean population there are nine common β-gene haplotypes, each with a different associated β-thalassemic mutation. The different haplotypes and molecular variants of β thalassemia vary widely in frequency, although their distribution among β thalassemic patients is largely similar to that in the nonaffected members of the population from which these patients were derived. The reasons for these curious

disclosed yet another pattern of molecular defects. These conditions, which are classified into $^{G}\gamma$ and $^{G}\gamma^{A}\gamma$ forms depending on the structure of the hemoglobin F that is produced, involve deletions of different sizes of the ϵ–$^{G}\gamma$–$^{A}\gamma$–δ–β globin-gene cluster on the short arm of chromosome 11 (figure). In $^{G}\gamma^{A}\gamma\delta\beta$ thalassemia the deletions start 3′ to the β gene and remove the whole of the β and part of the δ genes, but leave the 5′ end of the δ gene intact (Bernards et al., Gene *6,* 265–280, 1979; Ottolenghi et al., Nature *278,* 654–657, 1979). Two deletions responsible for $^{G}\gamma^{A}\gamma$ HPFH extend farther in the 5′ direction and remove all of the δ gene, one finishing in the region of the Alu repeat sequence, which lies 5′ to the δ locus, while another extends 5 kb farther upstream (Mears et al., PNAS *75,* 1222–1226, 1978; Bernards and Flavell, NAR *8,* 1521–1534, 1980; Tuan et al., Nature *285,* 335–337, 1980). Both $\delta\beta$ thalassemia and HPFH are characterized by the production of considerable amounts of γ chains in adult life; relatively more are synthesized in HPFH than in $\delta\beta$ thalassemia. Hence it has been suggested that the Alu repeat region may be involved in the regulation of γ-chain synthesis, and that its removal is required for the HPFH phenotype, but this notion is based only on a few detailed structural analyses, most of which are incomplete at the 3′ end of the deletions. There is no knowledge of what the effects of juxtaposition of normally distal chromosomal sequences may be on the activity of the β-gene cluster.

The difficulties of interpreting the functional effects of these deletions are exemplified further by the different lesions that result in $^{G}\gamma\delta\beta$ thalassemia. One lesion involves an inversion of most of the region between the $^{A}\gamma$- and δ-globin genes together with two deletions, a novel rearrangement that may have resulted from a rare, double-intrachromosomal crossing-over event (Jones et al., Nature *291,* 39–44, 1981). Two other types of $^{G}\gamma\delta\beta$ thalassemia are due to deletions involving the δ and β genes and either part or all of the $^{A}\gamma$ gene (Orkin et al., J. Clin. Invest. *64,* 866–869, 1979; Jones et al., NAR *9,* 6813–6825, 1981). All three forms have similar phenotypes, yet in the first type the Alu repeat region is present, though in a reverse orientation, while in the latter two cases it has been deleted. It may be, of course, that a single remaining $^{G}\gamma$ locus, even if fully active, cannot compensate for the lack of β-chain production; if it could, these disorders might have had an HPFH phenotype.

Although these findings do not argue conclusively for or against there being regions involved in the regulation of the changes in globin-gene expression, it is apparent that most of the major downstream deletions are associated with persistent activity of upstream loci that are normally inactivated during development. Further evidence for the interdependence of the loci within this gene cluster comes from studies of one form of $\gamma\delta\beta$ thalassemia that results from a long deletion involving both γ-globin genes; although the β gene is intact, β-globin synthesis is markedly reduced or possibly even abolished (van der Ploeg et al., Nature *283,* 637–642, 1980). This has led to the notion that the γ and β genes lie in functionally distinct domains, the activities of which are mutually exclusive (Bernards and Flavell, op. cit.). Despite the lack of evidence for specific regulatory sequences, it does appear that persistent γ-chain synthesis in these conditions is related in some way to *loss* of extensive regions of the γ–δ–β globin-gene cluster, and is not simply due to any sort of major rearrangement. Thus a duplication of the $^{G}\gamma$ gene resulting in effect in a 5 kb insertion and giving the arrangement $^{G}\gamma$–$^{G}\gamma$–$^{A}\gamma$ has recently been observed (Trent et al., NAR *9,* 6723–6733, 1981). The phenotypic effects of this insertion, as for the triplicated α genes, are minimal.

It is becoming clear that the molecular basis for most of the thalassemias is a simple cis-acting mutation within the globin-gene cluster, and that these disorders are extraordinarily diverse; often, apparently homozygous individuals are in fact compound heterozygotes for different molecular forms of thalassemia. The coinheritance of one or more deletion or nondeletion forms of α thalassemia can significantly ameliorate homozygous β^{+} or β^{0} thalassemia (Weatherall et al., Lancet *1,* 527–529, 1981). Other human genetic disorders will probably show similar molecular diversity; there is no reason to believe that thalassemia is unique in this respect.

Further analysis of some of the splicing defects that give rise to the β^{+} thalassemias, with more refined in vitro transcription systems, will provide useful information about the details of processing of the primary transcript. Although no mutations affecting transcription of the globin genes have yet turned up, there is still hope; some of the nondeletion α thalassemias and the mild β^{+} thalassemias are possible candidates. With regard to the developmental genetics of hemoglobin, we may have concentrated on the wrong conditions; the forms of pancellular HPFH and $\delta\beta$ thalassemia studied so far are associated with extensive gene deletions, and may not be the best models for analysis of gene regulation during normal development. Conditions such as nondeletion HPFH, although less dramatic in their phenotypic effects, may reflect more closely the changes that take place during normal fetal development. It is encouraging that the genetic determinants for some of these conditions have been shown to be linked to restriction enzyme polymorphisms within and flanking the γ–δ–β globin-gene cluster, and their locations may soon be pinpointed more precisely.