1 Haemoglobin and the genetics of haemoglobin synthesis

Haemoglobins and their structure and function

The haemoglobin molecule contained within red blood cells is essential for human life, being the means by which oxygen is transported to the tissues. Other functions include the transport of carbon dioxide (CO_2) and a buffering action (reduction of the changes in pH that would otherwise be expected when an acid or an alkali enters or is generated in a red cell). A normal haemoglobin molecule has a molecular weight of 64–64.5 kDa and is composed of two dissimilar pairs of polypeptide chains, each of which encloses an iron-containing porphyrin designated haem (Fig. 1.1). Haem is essential for oxygen transport while globin serves to protect haem from oxidation, renders the molecule soluble and permits variation in oxygen affinity. The structure of the haemoglobin molecule produces an internal environment of hydrophobic radicals, which protects the iron of haem from water and thus from oxidation. External radicals are hydrophilic and thus render the haemoglobin molecule soluble. Both haem and globin are subject to modifications. The iron of haemoglobin is normally in the ferrous form (Fe^{2+}). Haem is able to combine reversibly with oxygen so that haemoglobin can function as an oxygen-transporting protein. Oxidation of iron to the ferric form (Fe^{3+}) is a less readily reversible reaction, converting haem to haematin and haemoglobin to methaemoglobin, a form of haemoglobin that cannot transport oxygen. Auto-oxidation of haemoglobin to methaemoglobin is a normal process. About 3% of haemoglobin undergoes this process each day with about 1% (0.4–1% in one study) of haemoglobin being methaemoglobin [1, 2]. Methaemoglobin is converted back to haemoglobin mainly by the action of NADH-cytochrome b5-methaemoglobin reductase.

The haemoglobin molecule can also combine with CO_2, haemoglobin being responsible for about 10% of its transport from the tissues to the lungs; transport is by reversible carbamation of the N-terminal groups of the α chains of haemoglobin. Because carbamated haemoglobin has a lower oxygen affinity than the non-carbamated form, binding of the CO_2 produced by the metabolic processes in tissues facilitates oxygen delivery to tissues. In addition, non-oxygenated haemoglobin can carry more CO_2 than oxygenated haemoglobin so that unloading of oxygen to the tissues facilities the uptake and transport of CO_2. Because of its buffering action (mopping up of protons, H^+), haemoglobin also contributes to keeping CO_2 in the soluble bicarbonate form and thus transportable. The reaction $CO_2 + H_2O \rightarrow HCO_3^- + H^+$ is facilitated.

Haemoglobin also has a role in nitric oxide (NO) transport and metabolism. Haemoglobin is both a scavenger of nitric oxide and an active transporter. Nitric oxide is produced in endothelial cells and neutrophils by the action of nitric oxide synthases [2–5]. It has a very high affinity for oxyhaemoglobin so that blood levels are a balance between production and removal by binding to oxyhaemoglobin. Nitric oxide is a potent vasodilator, but this effect is limited by its binding to haemoglobin. The iron atom of a haem group of oxyhaemoglobin (preferentially the haem enclosed in the haem pocket of an

Haemoglobinopathy Diagnosis, Fourth Edition. Barbara J. Bain and David C. Rees.

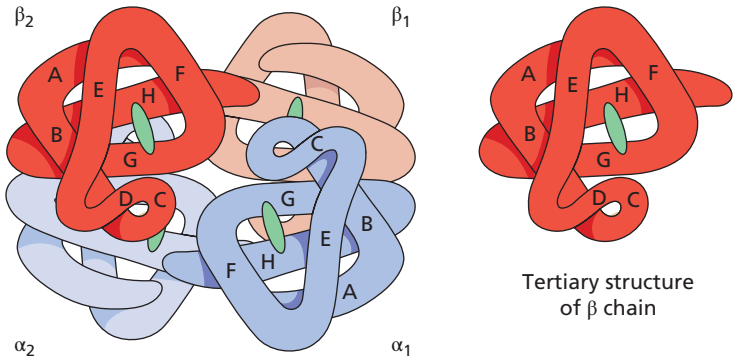

Fig. 1.1 Diagrammatic representation of the tertiary structure of a haemoglobin monomer (a β globin chain containing a haem group) and the quaternary structure of haemoglobin; upper case letters indicate homologous α helixes.

α chain), binds nitric oxide. A haemoglobin molecule with nitric oxide bound to two haem groups strikingly favours the deoxy conformation so oxygen is readily released. Nitric oxide-haemoglobin is subsequently converted to methaemoglobin with release of nitric oxide and production of nitrate ions, which are excreted. Since deoxyhaemoglobin has a much lower affinity for nitric oxide, hypoxic conditions could leave more nitric oxide free and lead to vasodilation, which is of potential physiological benefit. In addition, deoxyhaemoglobin can convert nitrite to nitric oxide, again favouring vasodilation.

Nitric oxide also causes *S*-nitrosylation of a conserved cysteine residue (Cys^{93}, E15) of the β globin chain of oxyhaemoglobin to form *S*-nitrosohaemoglobin. This occurs in the lungs. In this circumstance, the bioactivity of nitric oxide may be retained with nitric oxide being delivered to low molecular weight thiol-containing molecules to reach target cells such as the smooth muscle of blood vessels. Oxygenation of haemoglobin favours *S*-nitrosylation. Conversely, deoxygenation favours release of nitric oxide. This may be an important physiological process with nitric oxide being released in peripheral tissues where it can facilitate arteriolar dilation. The oxy form of *S*-nitrosohaemoglobin is a vasoconstrictor whereas the deoxy form is a vasodilator. Lack of oxygen could thus again favour vasodilation.

In normal circumstances, the ability of haemoglobin to scavenge or destroy nitric oxide is reduced by the barrier to nitric oxide diffusion that is provided by the red cell membrane. However, in haemolytic anaemias with increased free plasma haemoglobin, binding and inactivation can be almost immediate, leading to impaired vascular responses to nitric oxide [5]; inactivation of nitric oxide by haemoglobin in the plasma may thus contribute to the pulmonary hypertension that can be a feature of sickle cell anaemia and also to the hypertension that has been observed with some haemoglobin-based blood substitutes.

Surprisingly, the α globin genes are expressed in endothelial cells with the α globin participating in nitric oxide scavenging [6]. Individuals with deletion of one or two α globin genes have enhanced nitric oxide-induced vasodilation [7].

As a result of the synthesis of different globin chains at different stages of life (Fig. 1.2) there is a difference in the type of haemoglobin present in red cells between adult life and the fetal and neonatal periods (Table 1.1, Fig. 1.3). In adults, 96–98% of haemoglobin is haemoglobin A (A = adult), which has two alpha (α) chains and two beta (β) chains. The name 'haemoglobin A' was given by Linus Pauling and colleagues in 1949 when they discovered that asymptomatic carriers of sickle cell disease had two different haemoglobins, which they designated haemoglobin A and haemoglobin S [8]. A minor

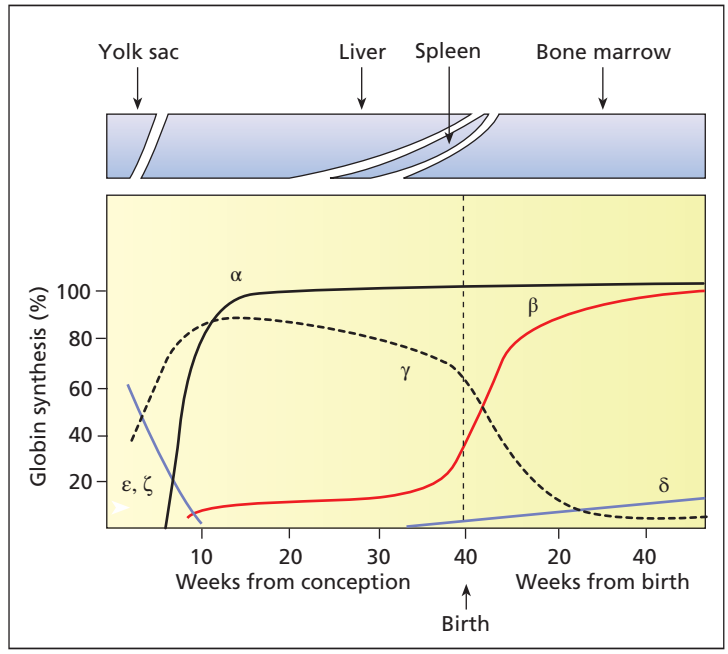

Fig. 1.2 Diagrammatic representation of the sites and rates of synthesis of different globin chains in the embryonic and fetal periods and during infancy.

Table 1.1 Haemoglobins normally present during adult, fetal and embryonic periods of life.

Haemoglobin species	Globin chains	Period when normally present
A	$\alpha_2\beta_2$*	Major haemoglobin in adult life
A_2	$\alpha_2\delta_2$	Minor haemoglobin in adult life; even more minor in late fetal and neonatal life
F	$\alpha_2{}^{G}\gamma_2$, $\alpha_2{}^{A}\gamma_2$ or $\alpha_2{}^{A}\gamma{}^{G}\gamma$	Minor haemoglobin in adult life, major haemoglobin in fetal life with a declining percentage through the neonatal period
Gower 1	$\zeta_2\varepsilon_2$	Significant haemoglobin during early intrauterine life
Gower 2	$\alpha_2\varepsilon_2$	Significant haemoglobin during early intrauterine life
Portland or Portland 1†	$\zeta_2\gamma_2$	Significant haemoglobin during early intrauterine life

* Can also be designated $\alpha^{A}{}_2\beta^{A}{}_2$ to distinguish the globin chains of haemoglobin A from those of variant haemoglobins.
† Haemoglobin Portland 2 ($\zeta_2\beta_2$) has been observed in α thalassaemia syndromes but is unlikely to occur in significant amounts during normal development.

haemoglobin, haemoglobin A_2, has two α chains and two delta (δ) chains. Its existence was first reported in 1955 by Kunkel and Wallenius [9]; they noted its increased level in thalassaemia minor and that it was reduced or absent in neonates. A very minor haemoglobin in adults but the major haemoglobin during fetal life and the early neonatal period is haemoglobin F or fetal haemoglobin, which has two α chains and two gamma (γ) chains. There are two species of haemoglobin F, designated $^{G}\gamma$ and $^{A}\gamma$, with glycine and alanine respectively at position 136 of the γ chain. In addition, the $^{A}\gamma$ chain shows polymorphism at position 75, which may be occupied by threonine rather than the more common isoleucine [10], a polymorphism that

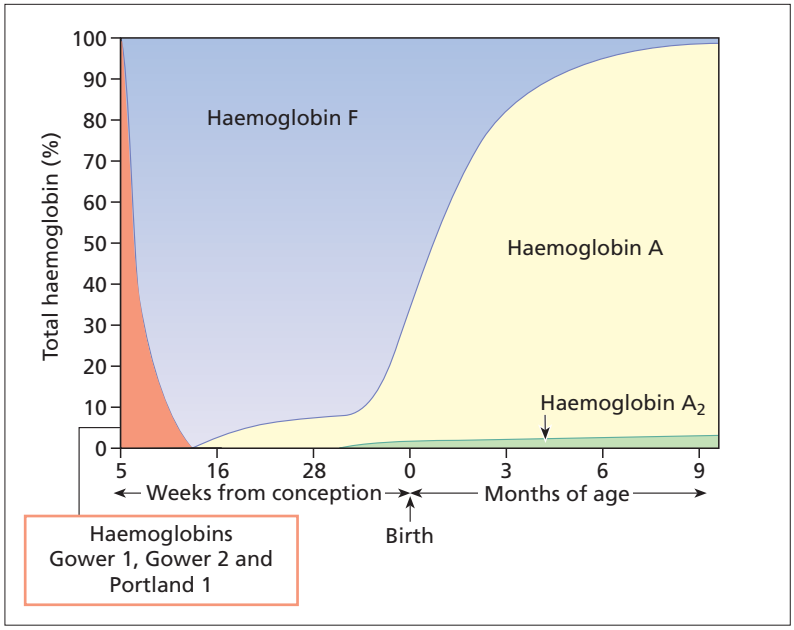

Fig. 1.3 Diagrammatic representation of the average percentages of various haemoglobins present in the embryonic and fetal periods and during infancy.

was previously referred to as haemoglobin F-Sardinia. In the early embryo, haemoglobin is synthesised in the yolk sac and specific embryonic haemoglobins are produced – Gower 1, Gower 2 and Portland (or Portland 1). They contain globin chains that are synthesised in significant amounts only during embryonic life, specifically zeta (ζ) and epsilon (ε) chains (see Table 1.1). Haemoglobins Gower 1 ($\zeta_2\varepsilon_2$) and Gower 2 ($\alpha_2\varepsilon_2$) were first described by Huehns and colleagues in 1961 [11], being named after Gower Street, in London, in which University College Hospital is situated. Portland 1 ($\zeta_2\gamma_2$) was described in 1967 and was so named because it was first identified in the University of Oregon in Portland, Oregon [12]. By five weeks of gestation, ζ and ε chains are already being synthesised in primitive erythroblasts in the yolk sac. From the sixth week onwards these same cells start to synthesise α, β and γ chains. Starting from about the 10th to the 12th week of gestation there is haemoglobin synthesis in the liver and the spleen with production of fetal and later adult haemoglobin. Production of the various embryonic, fetal and adult haemoglobins is synchronous in different sites. Later in intrauterine life, the bone marrow takes over as the main site of haemoglobin synthesis and increasing amounts of haemoglobin A are produced. In adult life, bone marrow erythroblasts synthesise haemoglobin A and the minor haemoglobins.

The embryonic haemoglobins have a higher oxygen affinity than haemoglobin A, similar to that of haemoglobin F [13]. They differ from haemoglobins A and F in that they continue to bind oxygen strongly, even in acidotic conditions [13]. In the case of Gower 2, impaired binding to 2,3-diphosphoglycerate (2,3-DPG) is the basis of the increased oxygen affinity [14].

Formation of the haemoglobin A molecule starts with formation of an αβ dimer. Normally α chains are produced in slight excess. Alpha haemoglobin stabilising protein (AHSP) acts as a molecular chaperone, facilitating formation of the dimer and preventing the precipitation of free α chains, which would lead to generation of reactive oxygen species with resultant damage to cells. Any free β chains are soluble.

Haemoglobin can undergo post-translational modifications (see also Chapter 6). Glycosylation occurs with formation of haemoglobins $A_{1a\text{-}e}$, but principally of haemoglobin A_{1c}. In normal individuals haemoglobin A_{1c} may constitute up to 4–6% of total haemoglobin but in patients with diabetes mellitus it can be much higher. It is also

increased in the acquired immune deficiency syndrome (AIDS) [15]. In individuals with a shortened red cell life span the percentage of haemoglobin A_{1c} is lower. Another minor fraction, formed on ageing, is haemoglobin A_{III}, in which glutathione is bound to the cysteine at β93. Unmodified haemoglobin can be distinguished by use of the designation haemoglobin A_0. In the fetus about 20% of haemoglobin F shows acetylation of the γ chain but this is not a major feature of other normal human globin chains [10]. Exposure to carbon monoxide, the product of incomplete combustion of hydrocarbons, leads to the formation of carboxyhaemoglobin. In normal individuals carboxyhaemoglobin comprises 0.2–0.8% of total haemoglobin but in heavy smokers it may be as much as 10–15%. Small amounts of sulphhaemoglobin (<0.5%) [1] and methaemoglobin are also formed in normal subjects. Methaemoglobin (see earlier) is usually less than 1% of total haemoglobin. Other post-translational modification of globin chains includes carbamylation, pyruvatisation and acetaldehyde adduct formation [16]. Glutathionylation is increased in diabetes mellitus [17] and by the administration of certain anti-epileptic drugs (phenobarbital and carbamazepine) [18]. Post-synthetic modification of a haemoglobin molecule can also occur as a consequence of a mutation in a globin gene; either the abnormal amino acid or an adjacent normal amino acid can undergo post-translational conversion to another amino acid (see later). In addition, some abnormal haemoglobins in which there is a mutation of N terminal amino acid are particularly prone to acetylation, which occurs co-translationally [19].

The structure of haemoglobin is highly complex and can be viewed at four levels.

1 The primary structure is the sequence of the amino acids in the polypeptide that constitutes the globin chain.

2 The secondary structure is the arrangement of the polypeptide globin chains into α helixes (stabilised by hydrogen bonds) separated by non-helical turns. In the case of the β globin chain there are eight α helixes, designated A to H, whereas the α globin chain lacks the D helix residues; 70–80% of the amino acid residues of haemoglobin form part of the helixes.

3 The tertiary structure is the arrangement of the coiled globin chain into a three-dimensional structure that has a surface haem-containing pocket between the E and F helixes; binding of haem between two specific histidine residues in the E and F helixes respectively (Fig. 1.4) is essential for maintaining the secondary and the tertiary structure of haemoglobin.

4 The quaternary structure is the relationship between the four globin chains, which is not fixed. The strong $\alpha_1\beta_1$ and $\alpha_2\beta_2$ bonds (dimeric bonds) hold the molecule together in a stable form while the $\alpha_1\beta_2$ and $\alpha_2\beta_1$ bonds (tetrameric bonds) both contribute to stability, albeit to a lesser extent than the dimeric bonds, and permit the chains to slide on each other and rotate; alteration in the quaternary structure of haemoglobin is responsible for the sigmoid oxygen dissociation curve, the Bohr effect and the variation of oxygen affinity consequent on interaction with 2,3-DPG (see later). Contacts between like chains, $\alpha_1\alpha_2$ and $\beta_1\beta_2$, are also of physiological significance.

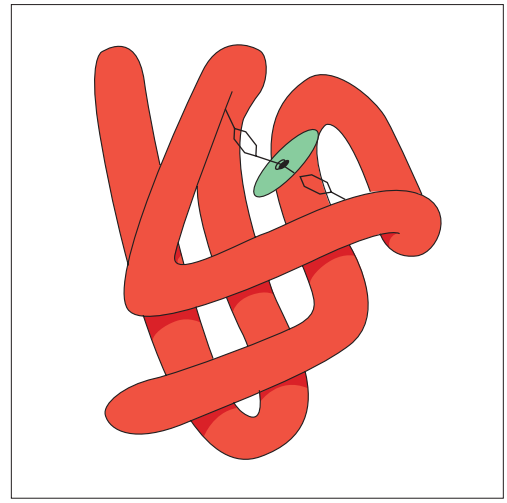

Fig. 1.4 Diagrammatic representation of a haemoglobin molecule with a haem group within the haem pocket, showing the relationship of the haem to two histidine residues of the globin chain, designated proximal and distal histidines; haem is bound to the proximal histidine while O_2 is bound to haem and to the distal histidine, both histidines being important for the integrity of the haem pocket.

The interaction between the four globin chains is such that oxygenation of one haem group alters the shape of the molecule in such a way that oxygenation of other haem groups becomes more likely. This is known as cooperativity and is reflected in the shape of the oxygen dissociation curve (Fig. 1.5). The cooperativity between the globin chains is shown diagrammatically in Fig. 1.6. It is consequent on the fact that in the deoxygenated state the Fe^{2+} atom is out of the plane of the porphyrin ring of haem. Oxygenation of Fe^{2+} causes it to move into the plane of the porphyrin ring and because of the link between haem and the histidine residues of globin there is an alteration in the tertiary structure of that haemoglobin monomer; this in turn causes the oxygenated monomer to alter its position in relation to other haemoglobin monomers, (i.e. the quaternary structure of the haemoglobin molecule is altered). The oxygenated haemoglobin molecule is smaller than the non-oxygenated molecule. Cooperativity between the globin chains is also the basis of the alkaline Bohr effect (often referred to simply as the Bohr effect) (i.e. the reduction of oxygen affinity that occurs when the pH falls from physiological levels of 7.35 to 7.45 towards 6.0). Increasing metabolism in tissues lowers the pH since there is increased production of CO_2 and of carbonic acid and, in addition, in anaerobic conditions there is generation of lactic acid. The Bohr effect therefore leads to enhanced delivery of oxygen to tissues such as exercising muscle. Similarly, the quaternary structure of haemoglobin makes possible the interaction of haemoglobin with 2,3-DPG, which enhances oxygen delivery. Synthesis of 2,3-DPG is increased by hypoxia. Marked anaemia can cause respiratory alkalosis, which enhances 2,3-DPG synthesis, thus compensating to some extent for the anaemia. There is also increased 2,3-DPG synthesis in renal failure, again partly compensating for the anaemia.

Oxygen affinity is reduced not only by acidosis and increased levels of 2,3-DPG but also by fever. All these effects are likely to be of physiological significance. Fever increases the

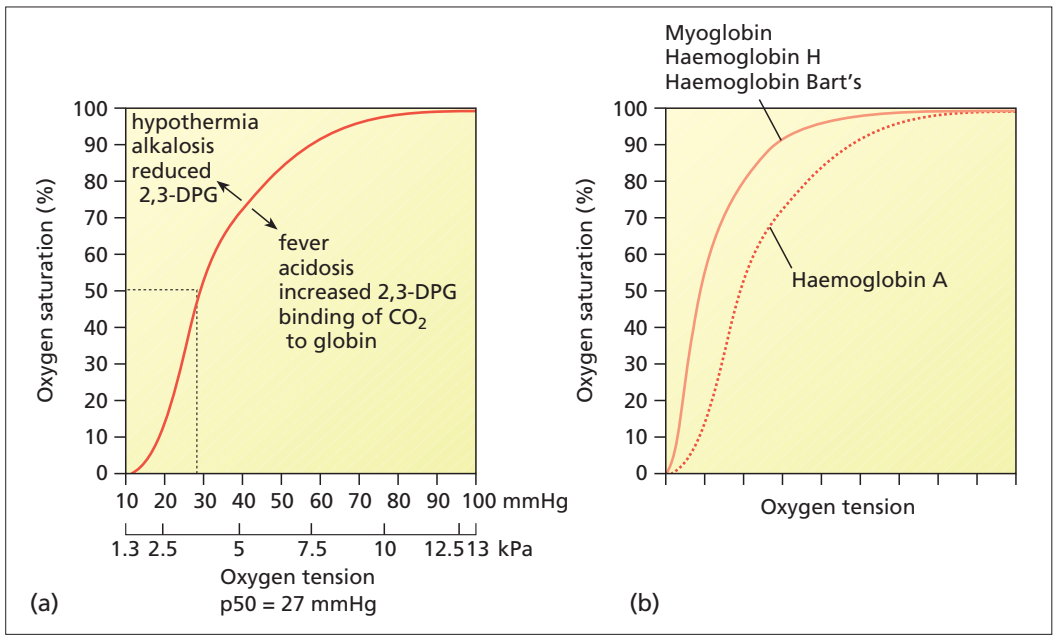

Fig. 1.5 Oxygen dissociation curve: (a) normal oxygen dissociation curve indicating the effects of alteration of pH, body temperature and 2,3-diphosphoglycerate (2,3-DPG) concentration on the oxygen affinity of haemoglobin; (b) a comparison of the hyperbolic oxygen dissociation curve characteristic of myoglobin and of abnormal haemoglobins that do not exhibit cooperativity with the sigmoid dissociation curve characteristic of haemoglobin A; haemoglobin A_2 has a dissociation curve similar to that of haemoglobin A but further to the right.

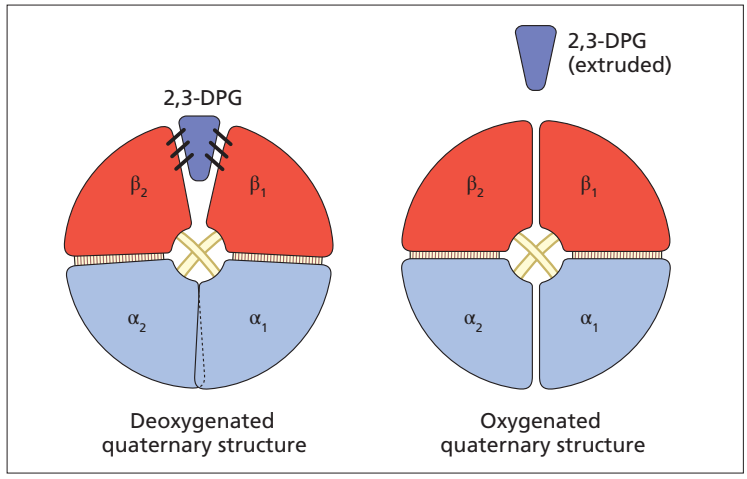

Fig. 1.6 Diagrammatic representation of the effect of oxygenation and deoxygenation on the quaternary structure of haemoglobin. The haemoglobin dimers ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) are stable, with the dimeric bonds between the α and the β chain having 34 contacts in both the deoxygenated and oxygenated forms. There are less strong $\alpha_2\beta_1$ and $\alpha_1\beta_2$ tetrameric bonds, with 17 contacts between the α chain and the β chain, in the deoxy form and a different 17 contacts in the oxy form. There are also $\alpha_1\alpha_2$ bonds with four inter-chain contacts in the deoxy form only. 2,3-DPG binds to the β chains (3 contacts with each chain) only in the deoxy form of the molecule. Oxygenation is associated with breaking and reforming of tetrameric ($\alpha_2\beta_1$ and $\alpha_1\beta_2$) contacts, breaking of $\alpha_1\alpha_2$ contacts, expulsion of 2,3-DPG and the assumption of a more compact form of the molecule. In the deoxygenated form the α chains are closer together and there is a cleft between the β chains whereas in the oxygenated form the α chains are further apart and the β cleft has disappeared.

metabolic rate so that decreased oxygen affinity, favouring offloading of O_2, is beneficial in this circumstance. The lower pH in tissues favours delivery of oxygen to sites of active metabolism, whereas the efflux of CO_2 in the lungs raises the pH and favours uptake of oxygen by haemoglobin. The oxygen dissociation curve is often right shifted, as a result of acidosis, in chronic renal failure; this ameliorates the effect of anaemia [20]. It will be noted that the acute effect of acidosis and the chronic effect of respiratory alkalosis both contribute to improved oxygen delivery to tissues.

Genetics of haemoglobin synthesis

Haem synthesis takes place in erythroid precursors from the proerythroblast stage to the reticulocyte stage. Eight enzymes, each under separate genetic control, are known to be necessary for haem synthesis [21]. Different stages of haem synthesis take place either in mitochondria or within the cytosol (Fig. 1.7). The first enzymatic reaction and the last three are in the mitochondrion whereas the four intermediate enzymatic reactions occur in the cytosol. The first rate-limiting step in haem synthesis is formation of δ-aminolaevulinic acid by condensation of glycine and succinyl CoA. This reaction is under the control of aminolaevulinate synthase (ala-synthase) with pyridoxal 5′-phosphate as a cofactor. In erythroid tissue the rate of formation of δ-aminolaevulinic acid is controlled by iron availability; iron deficiency causes iron regulatory proteins to bind to iron-responsive elements in the messenger ribonucleic acid (mRNA) for ala-synthase with resultant repression of translation. Synthesis of δ-amino laevulinic acid is followed by its entry into the cytosol where two molecules combine, under the influence of δ-aminolaevulinate dehydrase (ala-dehydrase), to form porphobilinogen. Four molecules of porphobilinogen in turn combine to form uroporphyrinogen III, which is then modified in two further steps to form coproporphyrinogen III. Coproporphyrinogen

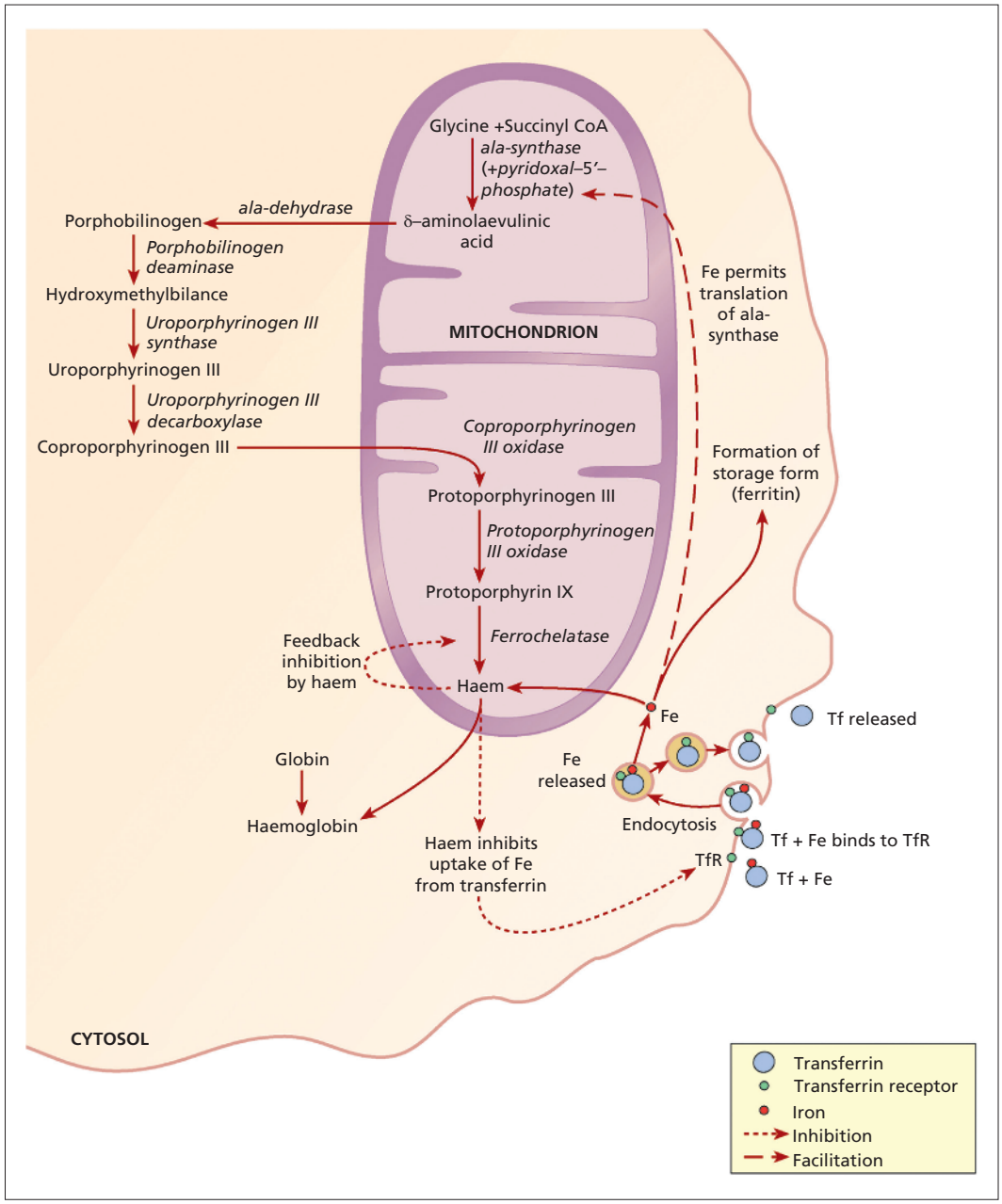

Fig. 1.7 Diagrammatic representation of haem synthesis. Tf, transferrin; TfR, transferrin receptor.

III enters the mitochondrion where it is converted to protoporphyrin IX. The final stage is the combination of ferrous (Fe^{2}+) iron with protoporphyrin IX to form haem, under the influence of ferrochelatase. Haem is also referred to as ferroprotoporphyrin.

Uptake of iron by erythroid cells is from transferrin (see Fig. 1.7). A molecule of transferrin with its attached iron first binds to a membrane transferrin receptor. The whole complex is internalised, in a process known as endocytosis. Iron is released from its carrier

within the endocytotic vesicle and, following reduction to the ferrous form, is either transferred to the mitochondrion for haem synthesis or is stored as ferritin within the cytoplasm. The transferrin molecule then detaches from the transferrin receptor and is released from the cell surface. There is negative feedback control of haem synthesis by haem, which both inhibits ferrochelatase and inhibits acquisition of iron from transferrin. Reduced cellular uptake of iron in turn inhibits production of δ-amino laevulinic acid. Uptake of iron by erythroid cells is enhanced by iron deficiency and by increased levels of erythropoietin. Both lead to combination of iron regulatory proteins with iron-responsive elements in the mRNA for the transferrin receptor protein. The mRNA is then protected from degradation, leading to increased expression of transferrin receptors on erythroid cell membranes and increased iron uptake.

Haem is necessary for normal folding of globin chains and prevents their precipitation [22]. Variant haemoglobins with impaired haem binding are unstable [22]. Haem is important in the regulation of globin chain synthesis. In haem-replete cells a protein known as haem-regulated inhibitor (HRI) is inactive, with the result that guanosine diphosphate (GDP) attached to an erythroid initiation factor, eIF2, is converted to guanosine triphosphate (GTP), leading to initiation of globin chain synthesis. When haem is deficient, HRI is activated by autophosphorylation and maintains eIF2–GDP in an inactive form so that globin chain translation is not initiated [22]. HRI is likely to be of relevance in β thalassaemia, with increased levels resulting from oxidative damage lessening the excess α chain synthesis.

Synthesis of α, β and γ globin chains takes place in erythroid precursors, from the pro-erythroblast onwards, and continues to the reticulocyte stage. Synthesis of δ chains ceases before the reticulocyte stage [23]. Haemoglobin A synthesis thus continues in reticulocytes, whereas synthesis of haemoglobin A_2 has been completed by the late erythroblast stage [24].

Globin chain synthesis takes place on ribosomes in the cytoplasm. Genes controlling globin chain synthesis are located in two clusters, on chromosomes 11 and 16 (Figs 1.8 and 1.9). The α gene cluster is close to the telomere of chromosome 16, at 16p13.3. The distance from the telomere shows polymorphic variation, from 170 to 430 kilobases (kb). The β gene is at 11p15.4. In addition to the functional globin genes these clusters contain 'pseudogenes', which are non-functional homologues of globin genes; they are transcribed but not translated. The α cluster of chromosome 16 extends over 28 kb and contains, in the following order: a ζ gene, *HBZ*, (also referred to as ζ_2); a pseudo-ζ gene (ψζ or ψζ1); two pseudo-α genes (ψα2 and ψα1); and two α

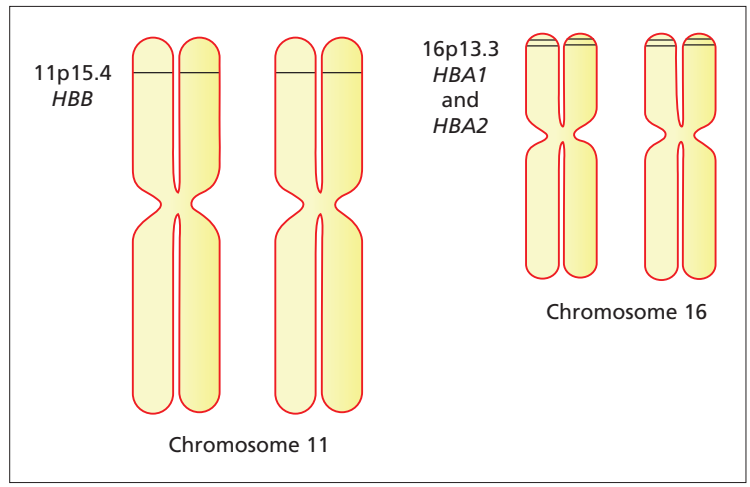

Fig. 1.8 Diagram of chromosomes 11 and 16 showing the positions of the β and α globin gene clusters.

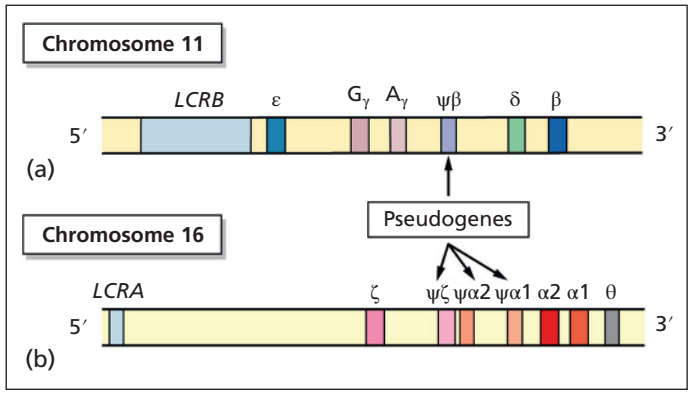

Fig. 1.9 Diagrammatic representation of the α and β globin gene clusters on chromosome 11 (a) and chromosome 16 (b).

genes, *HBA2* and *HBA1*, usually designated α2 and α1. The β cluster on chromosome 11 contains, in the following order: an ε gene, *HBE1*; two γ genes, *HBG2* and *HBG1*, usually designated $^{G}\gamma$ and $^{A}\gamma$ respectively; a pseudo-β gene (ψβ); a δ gene, *HBD*; and a β gene, *HBB*. There is wide variability of the α and β globin gene clusters between individuals and groups with duplications and triplications of ζ, ψζ and α being quite common. The overall structure of the two clusters are remarkably conserved amongst vertebrates and this has led to the hypothesis that all the globin genes, as well as the gene for the unlinked but related protein, myoglobin, arose from a common ancestor by the processes of duplication, unequal crossing over and sequence divergence. Many primitive invertebrates have only a single globin gene whereas fish and amphibians have an α and a β gene on the same chromosome. Birds have α and β genes on different chromosomes. All the human globin genes have three coding sequences (exons) and two intervening non-coding sequences (intervening sequences or introns) and are flanked by 5′ and 3′ non-coding sequences (referred to as untranslated regions, UTRs) (Fig. 1.10). The two α genes differ in structure in intron 2 and the 3′ UTR but the coding sequences are identical. As for all genes, coding is by means of triplets of nucleotides, known as codons, which code for a specific amino acid. 5′ to each gene is the promoter, promoters being sequences that bind ribonucleic acid (RNA) polymerase and transcription factors and are necessary for the initiation of transcription. Globin gene promoters share several conserved deoxyribonucleic acid (DNA) sequences that bind crucial transcription factors [25, 26]. These are summarised in Table 1.2.

The process by which globin chains are synthesised is shown diagrammatically in Fig. 1.10. Transcription is the process by which RNA is synthesised from a DNA template by the action of RNA polymerase. The entire globin gene, including the introns and the 5′ and 3′ UTRs, is transcribed. Transcription is controlled by interaction between the genes and transcription factors that bind both to promoters and to upstream regulatory elements referred to as the β-locus control region (*LCRB*) for the β cluster and the α-locus control region (*LCRA*) for the α cluster. The *LCRA* has four regulatory elements, DNase sites HS –48, HS –40, HS –33 and HS –10, also designated R1, R2, R3 and R4, of which HS –40 (R2) is the major regulatory element. It has been estimated that these enhancer elements contribute ~10%, 90%, <2–3% and < 2–3% respectively [27]. The *LCRB* includes five erythroid-specific DNase sites designated HS1, HS2, HS3, HS4 and HS5 of which HS3 is probably the most important in opening the chromatin structure to permit access of transcription factors and HS2 is probably the most important in enhancing globin chain synthesis [28]. There are also enhancers and facilitators [29] within introns of genes and downstream of the β and $^{A}\gamma$ genes. *Trans*-acting factors, encoded by genes on chromosomes other than 11 and 16, are vital for

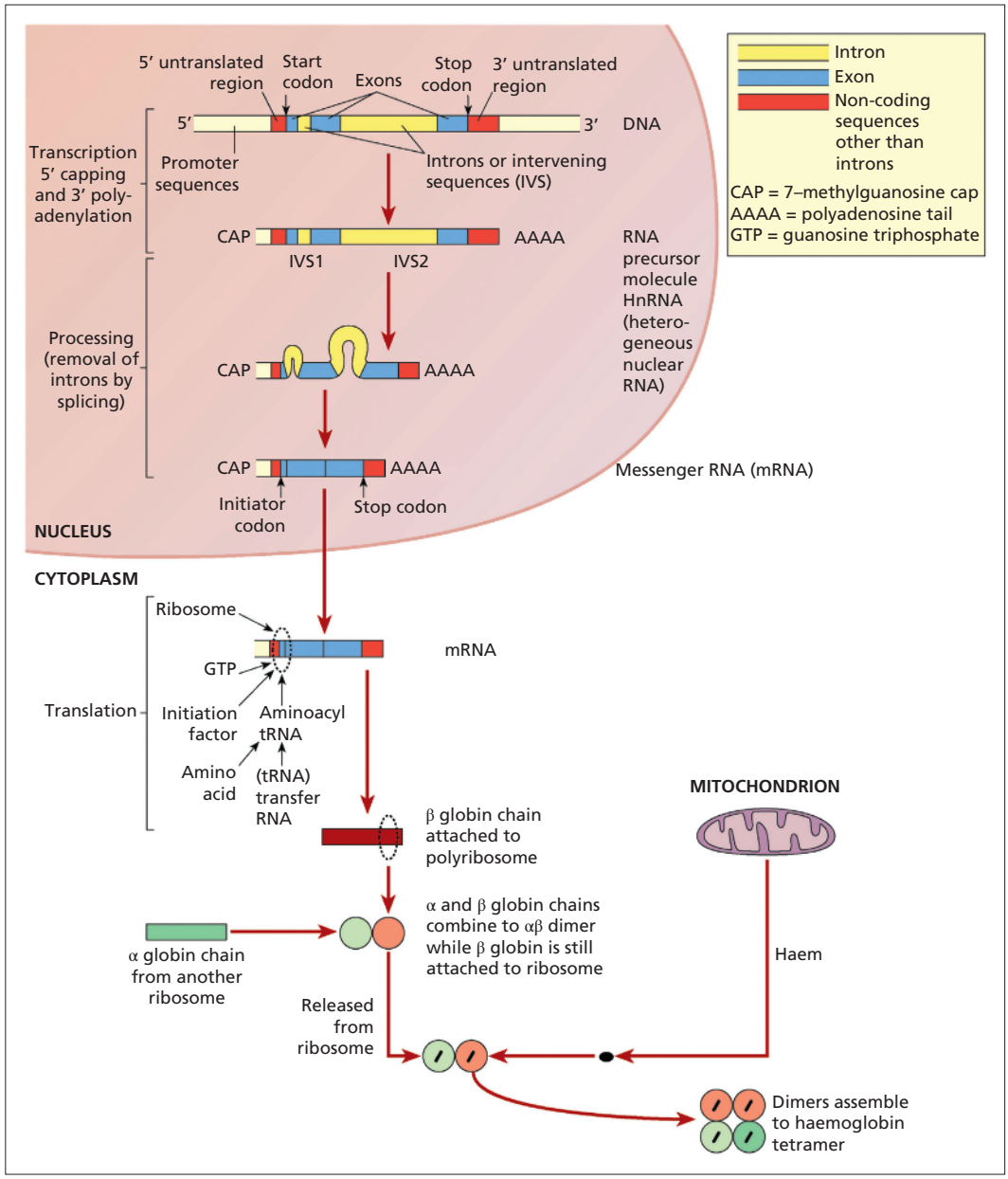

Fig. 1.10 Diagrammatic representation of ribonucleic acid (RNA) synthesis and processing and β globin chain synthesis. Although processes are shown sequentially, capping starts soon after transcription has started and therefore contemporaneously with transcription whereas polyadenylation necessarily occurs after completion of transcription.

the expression of globin genes. Relatively erythroid-specific *trans*-activating factors, including GATA1, ZFPM1 (previously known as FOG1) (which interacts with GATA1 in erythroid and megakaryocytic development), NFE2, KLF1 (previously known as EKLF), KLF2, NFE4 (SSP), Nrf-1, NFE2L2 (Nrf-2) and NFE2L1 (LCR-F1), contribute to regulation of

Table 1.2 The sequences showing CACCC, CCAAT and TATA homology in the promoters of globin genes; identical sequences in different genes are shown in bold red.

Gene	CACCC homology box	CCAAT homology box	TATA homology box
ζ (*HBZ*)		**CCAAT**	T**TATA**AAC
α1 and α2 (*HBA1* and *HBA2*)		**CCAAT**	C**ATA**AAC
ε (*HBE1*)		**CCAAT**	A**ATA**AAG
^{G}γ and ^{A}γ (*HBG2* and *HBG1*)	**CACCC**	**CCAAT/CCAAT**	A**ATA**AAA
β (*HBB*)	**CACCC**	**CCAAT**	C**ATA**AAA
δ (*HBD*)		**CCAA**C	C**ATA**AAA

gene expression by interacting either with the locus control regions or with the globin gene promoters to increase gene expression [30, 31]. These transcription factors interact, together with many other unidentified factors, in a complex and only partly understood way. KLF1 (Krüppel-like factor 1) is an enhancer of β chain synthesis and a repressor of γ chain synthesis, by means of its activation of BCL11A. BCL11A is part of a repressor complex, also including GATA1 and histone deacetylase 1 (HDAC1), which binds to a region near the 5′ end of the δ-globin gene, repressing the gene; this interaction is critical in the fetal-to-adult haemoglobin switch and in γ-globin gene silencing in adults [32]. Heterozygous inactivating mutations of *KLF1* can lead to an increased percentage of haemoglobin F [33] as can haploinsufficiency or downregulation of *BCL11A* [34]. Inactivating mutation of *KLF1* can also lead to an increase of haemoglobin A_2 [35] together with other red cell abnormalities including microcytosis and pyruvate kinase deficiency. SSP (stage selector protein) is an enhancer of δ and γ chain synthesis [30]. NFE4p22 is an enhancer of Gγ and Aγ genes [36]. FKLF and FKLF2 are enhancers of the embryonic ε gene and the γ genes [36, 37]. *SUPT5H* encodes a putative transactivating factor and mutation can lead to β thalassaemia [38]. In addition to transcription factors that are relatively specific to erythroid cells, globin gene expression is also influenced by general transcription factors including AP-1 (subunits encoded by various genes), Sp1, YY1, USF1 (USF) and TAL-1 (SCL) [28, 30, 31]. Expression of the genes of the β cluster is also influenced by histone acetylases, which increase expression [36]. The Gγ and Aγ genes are repressed by histone deacetylases and histone deacetylase inhibitors such as butyrate upregulate γ gene expression [36]. Methylation of genes reduces expression and thus the demethylating action of azacitidine may be the mechanism by which it upregulates γ gene expression [36].

Nascent RNA molecules resulting from transcription are large, unstable and modified in the nucleus. Initially the 5′ end acquires a 7-methyl guanosine cap (CAP), which is probably added early during transcription, protects the 5′ end of the molecule from degradation and is required for initiation of translation; during this 'capping' process methylation of adjacent ribose residues also occurs. Following the completion of transcription, the majority of transcripts acquire a 3′ polyadenosine tail with the addition of 75 to several hundred adenylate residues. There is an AAUAA sequence near the 3′ end (within the 3′ UTR), which serves as a signal for 3′ cleaving of the transcript and polyadenylation. The polyadenylate tail is important for mRNA stability, provides a signal for transfer of mRNA from the nucleus to the cytoplasm and probably enhances translation. Finally the introns are excised to give a functional mRNA, which in most cases contains a single continuous open reading frame (ORF), encoding the sequence of the relevant protein, flanked by 5′ and 3′ UTRs.

Molecules of mRNA move from the nucleus to the cytoplasm where they bind to ribosomes and

serve as templates for the assembly of the polypeptide sequences of the globin chain. Each nucleotide triplet serves as a template for a specific amino acid that is covalently bound to and transported to the ribosome by transfer RNA (tRNA). tRNAs are specific both for a nucleotide triplet and for an amino acid. Amino acids are thus assembled in the correct sequence, and are covalently bound to each other by ribosomal enzymes, forming a polypeptide. This process is known as translation. An initiation codon, AUG, is essential for the initiation of translation; it is the first codon after the 5′ untranslated region and encodes methionine. Initiation requires the amino acid methionine, tRNA specific for methionine, GTP and an initiation factor. When the nascent molecule reaches 20–30 amino acid residues, the methionine is removed through the action of methionine aminopeptidase; this process is interfered with when mutation leads to the presence of certain residues in position 1 or even position 2 of the globin chain [39]. When the chain reaches 40–50 residues co-translational acetylation of the N-terminal residue can occur through the action of several acetyl transferases [40]. Whether this occurs to any great extent depends on the nature of the N-terminal residue. Thus the glycine of the γ chain is 10–15% acetylated whereas the valine of normal α, β and δ chains is resistant to acetylation. Rarely an aberrant amino acid residue present as a result of mutation leads to increased acetylation, as is the case in haemoglobin Raleigh [39]. There are 64 possible nucleotide triplets or codons, 61 of which encode amino acids (20 in all) and three of which do not; the latter serve as STOP or termination codons, leading to termination of globin chain synthesis. Transcription thus continues until a termination codon, UAA, UAG or UGA, is encountered. The termination codon is followed by the 3′ UTR.

The rate-limiting step of globin chain translation is the commencement of elongation, which is the next step after initiation. Transcription from the two α genes is equal up to the 8th week of gestation but thereafter the α2 gene becomes dominant and, in adult life, the ratio of α2 to α1 mRNA is 2.6–2.8:1 [41]. Translational efficiency differs somewhat so that the α2 gene directs synthesis of about twice as much α chain as the α1 gene. There is more α than β mRNA, probably about two and a half times as much, but β chain synthesis is more translationally efficient than α chain synthesis and α chains are therefore produced only slightly in excess of β chains [41]. Control of globin chain synthesis is probably mainly at the level of transcription with translational control being less important. Translation is dependent on the presence of haem. In iron deficiency, reduced availability of haem leads to inactivation of the initiation factor and thus reduced synthesis of globin chains. The α and β globin chains are synthesised on different polyribosomes. Combination of free α chain with β chain that is still attached to the polyribosome, to form an αβ dimer, may contribute to release of the β chain from the ribosome. Incorporation of haem probably occurs after release from the polyribosome.

Globin mRNA is unusually stable so that translation can continue for up to 3 days after cessation of transcription, and approximately 20% of the total amount of erythrocyte globin is synthesised in the anucleate reticulocyte after its release from the bone marrow. Both the α and β globin genes have structural determinants in their 3′ UTRs that are important for mRNA stability [30].

Normal haemoglobins

The normal haemoglobins beyond the neonatal period are haemoglobin A and two minor haemoglobins, haemoglobin A_2 and haemoglobin F.

Haemoglobin A_2

In haematologically normal adults, haemoglobin A_2 comprises about 2–3.5% of total haemoglobin. The percentage is much lower at birth, about 0.2 to 0.3% or even lower, with a rise to adult levels during the first 2 years of life. The steepest rise occurs in the first year but there is a continuing slow rise up to 3 years of age [42] (Fig. 1.11). In the normal adult population, the percentage of haemoglobin A_2 shows a Gaussian distribution. It has functional properties that are

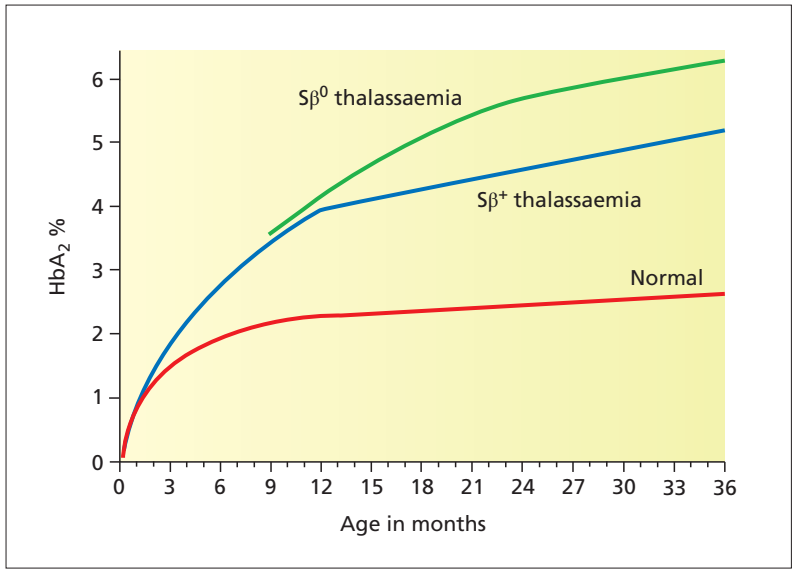

Fig. 1.11 Diagram showing the rate of rise of haemoglobin A_2 in haematologically normal Jamaican babies and in babies with sickle cell/β thalassaemia. (Adapted from reference [37].)

very similar to those of haemoglobin A [23] (similar cooperativity and interaction with 2,3-DPG) although, in comparison with haemoglobin A, it inhibits polymerisation of haemoglobin S [43] and has a higher oxygen affinity [14]. It has a pancellular distribution.

The reduced rate of synthesis of haemoglobin A_2, in comparison with haemoglobin A, reflects the much slower rate of synthesis of the δ chain in comparison with the β chain. This in turn appears to be consequent in part on a reduced rate of transcription of δ mRNA caused by a difference in the promoter region of these two genes; the δ gene has a CCAAC box rather than the CCAAT box of the β gene [23] and in addition lacks the CACCC sequence that is present in the β promoter (see Table 1.2). In addition, γ chain messenger RNA is unstable and there is also an influence of sequences in IVS2 [44]. The haemoglobin A_2 percentage in adults is controlled by two different genetic mechanisms [45]. Analysis of single nucleotide polymorphisms (SNPs) shows that alleles in the region of the *HBS1L* and *MYB* loci at 6q23.3 influence both the percentage of F cells and the A_2 percentage while alleles around the *HBB* locus at 11p15.4 influence the percentage of haemoglobin A_2 [45]; however, it does not appear to be the *HBS1L* gene itself that is responsible [46]. Inactivating mutations in *KLF1* can cause a borderline to moderate increase in haemoglobin A_2, up to 3.5–4.7%, with red cell indices typical of β thalassaemia heterozygosity [34, 35, 47, 48]. The proportion of haemoglobin A_2 is slightly reduced by absolute or functional iron deficiency (see Table 6.3) and by α, δ and δβ thalassaemia. In γδβ thalassaemia, the rate of synthesis, but not the proportion of haemoglobin A_2, is reduced since synthesis of γ and β chains is reduced, as well as δ chain synthesis. The proportion of haemoglobin A_2 is increased in the great majority of patients with β thalassaemia trait, in the majority of patient with haemoglobin E heterozygosity [49] and in some patients with an unstable haemoglobin.

Many δ chain variants and δ thalassaemias occur, at least 90 and 25 respectively having been described; some of the δ chain variants are thalassaemic variants. Around 1–2% of individuals of African ancestry have the variant haemoglobin designated haemoglobin A_2' (A_2 prime) or haemoglobin B_2 ($\delta^{16Gly\rightarrow Arg}$). In one study, it was the most common haemoglobin variant detected after haemoglobins S and C [44]. Haemoglobin A_2' is readily detected on high performance liquid chromatography (HPLC) (Fig. 1.12) and isoelectric focusing, and may also be detected on capillary electrophoresis (see Fig. 2.27). The δ thalassaemias are also common in some ethnic groups, (present in 1% of Sardinians) [14]. Although some δ chain variants are unstable or have increased oxygen affinity, the low percentage

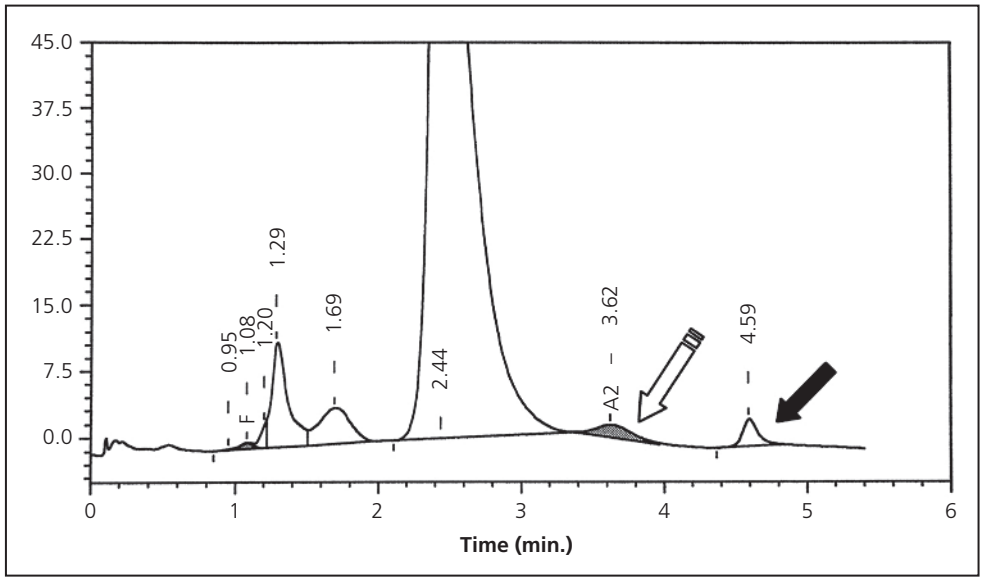

Fig. 1.12 High performance liquid chromatography (HPLC) chromatogram (Bio-Rad Variant II) showing a split haemoglobin A_2 resulting from heterozygosity for haemoglobin A_2'. The white arrow shows haemoglobin A_2 and the black arrow haemoglobin A_2'.

of haemoglobin A_2 means that δ thalassaemia and δ chain variants are of no clinical significance. However, their presence complicates the diagnosis of β thalassaemia trait (see page 141).

Haemoglobin F

Haemoglobin F is the major haemoglobin during fetal life. Its oxygen affinity is higher than that of haemoglobin A and this facilitates oxygen transfer from the mother to the fetus. However, it should be noted that fetal development can be normal in the offspring of mothers with very high levels of haemoglobin F [50]. Its oxygen dissociation curve is sigmoid. The increased oxygen affinity, in comparison with haemoglobin A, is attributable to its weak affinity for 2,3-DPG [10]. In comparison with haemoglobin A, haemoglobin F is less efficient at transporting CO_2. A significant proportion of haemoglobin F is acetylated.

During the first 2 years of life the haemoglobin F percentage falls progressively to values close to adult levels (Fig. 1.13) [51–53]. A slower fall to final adult levels may continue for several years, even up to puberty and beyond. The percentage of fetal haemoglobin present at birth is quite variable, usually being between 60% and 95%. During intrauterine life and at birth haemoglobin F shows a $^{G}\gamma$ to $^{A}\gamma$ ratio of approximately 2:1 to 3:1. Within the first few months of birth this changes to the adult ratio of approximately 2:3. In premature infants there is initially a plateau phase in haemoglobin F percentage lasting 20–60 days followed by a linear decrease similar to that in term babies [52]. At any given period after birth the spread of values is greater than in term babies. Initially there are more high values but after the first month of life values both higher and lower than those of term infants are observed [52]. Haemoglobin F levels at birth are higher in the babies of diabetic mothers and low-for-birth weight babies whereas in Down syndrome the switch to haemoglobin A is earlier [44].

In normal adults, haemoglobin F is heterogeneously distributed, being found in a subset of erythrocytes designated F cells. The proportion of F cells is closely correlated to the percentage of haemoglobin F and is highly variable. In one study the numbers of F cells ranged from 0.6% to 22% [54], although this varies depending on how the HbF is detected. The haemoglobin F percentage is determined by age, gender (slightly higher in women) and a number of

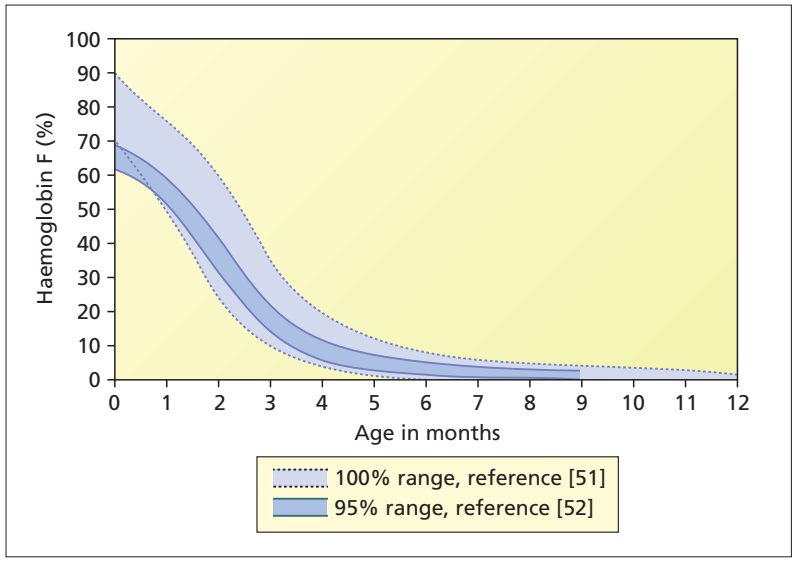

Fig. 1.13 The rate of fall of haemoglobin F percentage postnatally in normal and premature babies; the pale blue areas represent premature babies and the deep blue areas normal babies. (Adapted from references [51, 52].).

inherited characteristics both linked and unlinked to the β globin gene cluster. Variants in three main genes explain 20–50% of the variation in haemoglobin F levels and F cells seen in normal individuals and those with haemoglobinopathies. DNA sequences controlling the proportion of F cells and the percentage of haemoglobin F include [31, 51–58]:

- a polymorphism at position −158 of the Gγ gene (*HBG2*) (C→T being associated with a higher haemoglobin F) (also known as XmnI Gγ polymorphism), and largely explains the association between higher haemoglobin F percentage associated with the Senegal and Arab-Indian haplotype in sickle cell anaemia [59], β thalassaemia heterozygosity [60], haemoglobin E heterozygosity [61], haemoglobin E homozygosity [62] and inherited bone marrow failure syndromes [63];
- a *trans*-acting locus at 6q22.3–23.1 in the intergenic region between *HBS1L* and *MYB*, designated haemoglobin F quantitative trait locus 2 (*HBFQTL2*) or *HBS1L-MYB* intergenic polymorphism block 2 (*HMIP-2*), which affects F-cell numbers and haemoglobin F percentage in haematologically normal Europeans and in individuals with β thalassaemia major and sickle cell anaemia [59] and haemoglobin E homozygosity [62]; MYB is a transcription factor for a number of erythroid-specific genes and the intergenic sequences attenuate the function of nearby enhancers, reducing expression of *MYB* and upregulating haemoglobin F production [64];
- variants in the *BCL11A* gene at 2p16.1, designated haemoglobin F quantitative trait locus 5 (*HBFQTL5*), which affect the haemoglobin F percentage in normal Europeans, in individuals with β thalassaemia major and sickle cell anaemia [59], in β thalassaemia heterozygosity [60] and in haemoglobin E heterozygosity [61] and homozygosity [62]; BCL11A inhibits γ gene expression, this being of considerable therapeutic importance since gene therapy to downregulate expression can raise the haemoglobin F percentage sufficiently to ameliorate sickle cell anaemia and β thalassaemia.

Many other genetic variants have been identified which are also associated with natural variation in haemoglobin F levels or changes in haemoglobin F expression in experimental systems. These include:

- a *trans*-acting locus at Xp22.2, *FRMPD4* (*FCP1*, *HBFQTL3*), which affects F-cell numbers in normal individuals and in males with sickle cell anaemia [59, 65];
- a *trans*-acting locus on chromosome 8q, designated haemoglobin F quantitative trait locus 4 (*HBFQTL4*), which may be the *TOX* gene at 8q12.1 [66], which interacts with the common *HBG2* polymorphism (see earlier);
- variation of the number of repeats of a specific motif at −530 in the HS2 component of the β locus control region – $(AT)_XN_{12}GT(AT)_Y$;

• *SAR1A* at 10q22.1, which may influence haemoglobin F response to hydroxycarbamide [59];
• *KLF1* (previously known as *EKLF*) at 19p13.13, (*HBFQTL6*), reduced expression of which leads to reduced expression of the γ gene repressor, *BCL11A*, and therefore increased haemoglobin F; various inactivating heterozygous mutations have led to haemoglobin F levels of 1–3%, 1–7.4% and 3.3–19.5% and a dominant mutation can lead to haemoglobin F level of 40–50% with congenital dyserythropoietic anaemia [33, 67]; compound heterozygosity for S270X nonsense and K332Q mis-sense mutations led to haemoglobin F levels of 22–31%, whereas simple heterozygosity was not associated with an elevation [68];
• expression of *KBX3* (previously *CSDA*) at 12p13.2; cold shock protein domain A is a *trans*-acting suppressor of *HBG2* [69];
• *ZBTB7A* (*LRF*) at 19p13.3 inhibits γ gene expression in adult stage erythroid cells [70];
• *PPARGC1A* at 4p15.2 induces γ gene expression [70];
• the *HBBP1* pseudogene is a negative regulator in adult red cells [71];
• a *BGLT3* long non-encoding RNA is a positive regulator in fetal red cells [72];
• *ANTXR1* at 2p13.3 is associated with reduced haemoglobin F expression with several mutations being associated with a lower haemoglobin F in sickle cell anaemia associated with the Arab-Indian haplotype [73];
• *DNMT1* at 19p13.2,encodes a protein that represses γ gene expression, mutation sometimes leading to amelioration of β thalassaemia [74];
• *BMI1* at 10p12.2 encodes a protein of the polycomb repressor complex that represses haemoglobin F synthesis [75].

There is also a γ-globin gene silencer in a 3.5 kb region near the 5′ end of the δ-globin gene which, when deleted, can give rise to hereditary persistence of fetal haemoglobin [32].

The percentage of haemoglobin F is also affected by any increase in the number of γ genes.

The mechanism by which the polymorphisms in the *LCRB* at −530 bp to the Gγ gene influence γ chain synthesis appears to be that, in comparison with $(AT)_7T_7$, the $(AT)_9T_5$ sequence shows increased binding of EIF4EBP1 (BP-1), a negative *trans*-acting factor [76].

The distribution of haemoglobin F percentages in the population is skewed. In 85–90% of individuals, haemoglobin F is less than 0.6–0.7% and F cells are <4.5% [55, 57]. The other 10–15% of the population have values above these levels. The upper limit of normal is rather arbitrarily taken as 1%. It would probably be more accurate to take 0.6% or 0.7% as the upper limit of normal, excluding the 11% of males and 21% of females who have a slight elevation of the percentage of F cells and the haemoglobin F percentage as an X-linked dominant characteristic [55]. However, since the measurement of a low percentage of haemoglobin F is very imprecise, 1% is a practical upper limit.

Haemoglobin F is more markedly increased in patients with various inherited abnormalities of β globin chain synthesis, particularly the more severe forms of β thalassaemia and sickle cell disease, although these high levels are more related to the survival advantages of erythroid cells containing more haemoglobin F, rather than to direct stimulation of γ globin transcription (see Table 3.13). High levels are seen also in various acquired conditions (see Table 6.2).

Haemoglobin F can rise during pregnancy, usually not above 4% [77, 78].

Mutations in γ genes can lead not only to an increased percentage of haemoglobin F but also to haemoglobin F variants, at least 113 of which have been described.

Variant haemoglobins and abnormalities of globin chain synthesis

Nuclear DNA, including the DNA of globin genes, is subject to spontaneous mutation. This may be a point mutation (alteration of a single nucleotide) or a more extensive mutation, in which there is deletion, insertion or other alteration of more than one nucleotide. The types of mutation that can occur in globin genes are summarised in Table 1.3 [79–82]. In addition, expression of globin genes can be affected by DNA sequences outside the globin genes themselves, either enhancers acting in *cis* or genes on other chromosomes encoding *trans*-acting transcription factors (Tables 1.3 and 1.4 [34, 47, 48, 83–86]). The phenotype in

Table 1.3 Types of mutation that can occur in globin genes and adjoining sequences.

Type of mutation	Possible consequence	Examples
POINT MUTATIONS		
Within coding sequence, i.e. within an exon	**Same-sense or neutral mutation**, i.e. mutant codon codes for same amino acid as normal codon so there are no consequences	Many mutations are of this type; more than a third of theoretically possible point mutations would result in no alteration in the amino acid encoded
	Mis-sense mutation, i.e. mutant codon codes for a different amino acid from the normal codon; includes mis-sense mutations in which an abnormal amino acid interferes with the normal cleavage of the N-terminal methionine	Haemoglobin S, haemoglobin C, haemoglobin E; haemoglobin Marseille and haemoglobin South Florida (altered amino acid near N-terminus plus persisting methionine residue at the N-terminus of the β chain)
	Nonsense mutation, i.e. the mutant codon does not code for an amino acid and thus functions as a STOP or TERMINATION codon, producing a shortened globin chain	Haemoglobin McKees Rocks (two amino acids shorter than normal); α2 CD 116 GAG→TAG creating premature STOP codon and causing α thalassaemia
	New-sense mutation, i.e. conversion of a STOP codon to a coding sequence producing an elongated globin chain	Haemoglobin Constant Spring, haemoglobin Icaria, haemoglobin Seal Rock, haemoglobin Koya Dora, haemoglobin Paksé, haemoglobin Zunyi
	Mutation of the initiation codon	α or β thalassaemic disorder
	Gene conversion*	Conversion of $^{G}\gamma$ gene to $^{A}\gamma$ gene, giving $^{A}\gamma^{A}\gamma$ genotype
		Conversion of $^{A}\gamma$ gene to $^{G}\gamma$ gene, giving $^{G}\gamma^{G}\gamma$ genotype
		Conversion of ψζ1 to a gene that resembles ζ2 but is still non-functional
		Gene conversion between the α2 and α1 genes so that the same mutation is present in both, e.g. $\alpha 2^{Lys\rightarrow Glu}$ $\alpha 1^{Lys\rightarrow Glu}$, giving unusually high levels of haemoglobin I
		Gene conversion between a β gene and a δ gene leading to a haemoglobin A_2 variant (e.g. haemoglobin A_2 Flatbush)
	Gene conversion plus further point mutation*	Haemoglobin F Port Royal, resulting from a further point mutation in a $^{G}\gamma$-$^{G}\gamma$ gene complex
Within non-coding sequence, i.e. in an intron	Production of a new splice site leading to a structurally abnormal mRNA	Some β thalassaemias

Table 1.3 *Continued.*

Type of mutation	Possible consequence	Examples
Mutation 5′ or 3′ to the gene (i.e. outside the gene)	Mutation of an enhancer	Some β thalassaemias
	Reduced rate of synthesis of mRNA due to interference with 3′-end formation of mRNA	Some β thalassaemias
DELETION OR DUPLICATION OF ONE OR MORE GENES OR PSEUDOGENES (COPY NUMBER VARIANTS)		
D Deletion of one or more genes	Total loss of expression of relevant gene; occasionally also loss of function of an adjacent structurally normal gene	Most α thalassaemias, some β thalassaemias, δβ thalassaemias and γδβ thalassaemias; deletion of $^{G}\gamma$ gene ($-^{A}\gamma$), homozygosity for which causes anaemia and a reduced haemoglobin F percentage in the neonate; deletion of ψζ1
Deletion of genes with downstream enhancer being juxtaposed to remaining gene	Loss of β and δ gene function but enhanced function of remaining $^{G}\gamma$ ($\pm^{A}\gamma$) gene	Deletional hereditary persistence of fetal haemoglobin
Duplication of α gene	Triple, quadruple or quintuple α gene	ααα†/αα, αααα/αααα, ααααα‡/αα, ααααα/ααααα or αα/αααααα
Triplication of entire α globin gene cluster	Six α genes on a single chromosome	αα:αα:αα/αα
Duplication of $^{G}\gamma$ gene	Double, triple or quadruple $^{G}\gamma$ gene so that there are three, four or five γ genes on a chromosome	$^{G}\gamma^{G}\gamma^{A}\gamma$, $^{G}\gamma^{G}\gamma^{G}\gamma^{A}\gamma$ (homozygotes have been described with a total of 8 γ genes) or $^{G}\gamma^{G}\gamma^{G}\gamma^{G}\gamma^{A}\gamma$
Duplication of the ζ or ψζ gene	Double, triple or quadruple ζ/ψζ gene	ζ2ψζ1ψζ1/ζ2ψζ1 or ζ2ψζ1ψζ1/ζ2ψζ1ψζ1 or 4 ζ-like genes per chromosome
ABNORMAL CROSSOVER DURING MEIOSIS LEADING TO GENE FUSION		
α2-α1 fusion	Effective loss of one α gene but structurally normal α chain is encoded	$-\alpha^{3.7}$ thalassaemia
δβ fusion – simple crossover	Reduced rate of synthesis of structurally abnormal globin chain	Haemoglobin Lepore, e.g. haemoglobin Lepore-Washington/Boston, haemoglobin Lepore-Baltimore and haemoglobin Lepore Hollandia, or $\delta^{0}\beta^{+}$ thalassaemia [79]
δβδ fusion – double crossover with δ sequences on either side of β sequences	Reduced rate of synthesis of structurally abnormal globin chain	Haemoglobin Parchman
βδ fusion (with preservation of intact δ and β genes on either side of fusion gene, with or without additional mutation)		Anti-Lepore haemoglobins, e.g. haemoglobin Miyada, haemoglobin P-Nilotic, haemoglobin P-Congo haemoglobin Lincoln Park

(*Continued on pp. 20–21.*)

Table 1.3 *Continued.*

Type of mutation	Possible consequence	Examples
$^{A}\gamma\beta$ fusion	Synthesis of variant haemoglobin plus increased synthesis of haemoglobin F	Haemoglobin Kenya
$\beta^{A}\gamma$ fusion (with preservation of intact $^{G}\gamma$ and $^{A}\gamma$ genes and duplication of the δ gene)		Haemoglobin anti-Kenya
$^{G}\gamma\beta$ fusion	Variant was 37%, A_2 2.4%, F 6.6%, phenotype of β thalassaemia minor	Haemoglobin $^{G}\gamma$-β Ulan [80]
$^{G}\gamma^{A}\gamma$ fusion (designated -$^{G}\gamma^{A}\gamma$-)	Reduced rate of synthesis of haemoglobin F	γ thalassaemia
DELETION OF DNA SEQUENCES BUT WITHOUT A FRAME SHIFT IN CODING SEQUENCE		
Deletion of part of a coding sequence, either three nucleotides or a multiple of three	One to five amino acids missing but sequence otherwise normal	Haemoglobin Gun Hill (an unstable haemoglobin with five amino acids missing)
DELETION PLUS INVERSION		
Two deletions with inversion of intervening sequence	Deletion involving Aγ and δ plus β genes respectively but with preservation of an intervening region which is inverted	Indian type of deletional $^{A}\gamma\delta\beta^{0}$ thalassaemia
DELETION PLUS INSERTION		
Deletion with insertion of extraneous DNA between breakpoints	Same functional effect as deletion	One type of α^{0} thalassaemia, $--^{MED}$
INSERTION WITHIN A CODING SEQUENCE BUT WITHOUT A FRAME SHIFT		
Insertion of nucleotides, either three or multiples of three, e.g. by tandem duplication	Up to five extra amino acids	Haemoglobin Koriyama (an unstable haemoglobin with insertion of five codons in β gene, anti-Gun Hill); haemoglobin Grady (insertion of three codons in α gene)
INSERTION OF A DUPLICATED SEQUENCE INCLUDING THE START CODON		
Tandem duplication of codons flanking the start codon of the *HBB* gene	Reduced β chain synthesis	β^{+} thalassaemia; thalassaemia intermedia in a homozygote [81]
FRAME SHIFT MUTATIONS		
Alteration of the reading frame resulting from deletion, insertion, deletion plus insertion or deletion plus duplication	Abnormal amino acid sequence with an elongated globin chain (when a STOP codon is out of phase and translation continues until another 'in-frame' STOP codon is met); abnormal amino acid sequence with a truncated globin chain (when a premature STOP codon is created)	Haemoglobin Wayne (α chain), haemoglobin Tak (β chain), haemoglobin Cranston (β chain), some β thalassaemias including some dominant β thalassaemias, some α thalassaemias

Table 1.3 *Continued.*

Type of mutation	Possible consequence	Examples
CHROMOSOMAL TRANSLOCATION		
Unbalanced translocation (there may be a balanced translocation in a parent)	Extra α genes on a chromosome other than chromosome 16	Same significance as homozygous triplication of an α gene since there are a total of six α genes
	Loss of an α gene	α thalassaemia; may be part of a contiguous gene syndrome
DELETION OF A LOCUS CONTROL REGION		
Locus control region deleted, with or without deletion of relevant genes	Deletion of the locus control region of the β gene (*LCRB*)	$(\varepsilon)\gamma\delta\beta^0$ thalassaemia
	Deletion of the α gene enhancer (HS -40) 40 kb upstream of the ζ2 gene, also known as MCS-R2 (*LCRA*)	α^0 thalassaemia
CREATION OF A SEQUENCE THAT COMPETES FOR THE LOCUS CONTROL REGION		
Gain-of-function mutation within α gene cluster	Downregulation by competition with GATA1 for locus control region	α thalassaemia [82]

* Gene conversion is non-reciprocal genetic exchange between allelic or non-allelic homologous sequences so that one gene comes to resemble another more closely or becomes identical to it. This is responsible for maintaining the similarity between pairs of identical or similar genes.
† Either $\alpha\alpha\alpha^{anti3.7}$ or $\alpha\alpha\alpha^{anti4.2}$.
‡ Either $\alpha\alpha\alpha\alpha^{anti3.7}$ or $\alpha\alpha\alpha\alpha^{anti4.2}$.

Table 1.4 Mutations and polymorphisms occurring outside the globin gene clusters leading to abnormal globin chain synthesis [34, 47, 48, 83–86].

Mutation	Consequence
Mutation in *ATRX* gene at Xq21.1 which encodes a *trans*-acting factor regulating α gene expression	Haemoglobin H disease plus dysmorphism and severe learning difficulties
Mutation in *FCP1* (*HBFQTL3*) at Xp22.2	Hereditary persistence of fetal haemoglobin
An *HBS1L-MYB* intergenic polymorphism at 6q22.3–23.1	Hereditary persistence of fetal haemoglobin
Mutation in the *ERCC2* (*XPD*) gene at 19q13.22, which encodes one component of the general transcription factor, TFIIH [83]	Recessive trichothiodystrophy and β thalassaemia
Mutation in the *GATA1* gene at Xp11.23 [84]	X-linked thrombocytopenia and β thalassaemia (and, in one patient, congenital erythropoietic porphyria) [85]
BCL11A haploinsufficiency (2p16.1)	Increased haemoglobin F associated with cognitive impairment and facial dysmorphism [34]
KLF1 at 19p13.13	Increased haemoglobin A_2 [47, 48]
ASH1L at 1q22	β thalassaemia [86]

such cases can be that of thalassaemia trait or silent thalassaemia [87].

Variants of globin genes sometimes have no effect on the amino acid sequence. This occurs because, as mentioned earlier, there is redundancy in the genetic code, with a number of nucleotide triplets encoding the same amino acid. When a 'same sense' mutation occurs, the new codon resulting from the mutation codes for the same amino acid as the original codon and there is thus no effect on the final gene product. Similarly, mutation of a termination codon may be to a different termination codon. Many spontaneous variants of globin genes are 'same-sense' mutations. Point mutations can also result in a 'mis-sense' mutation when the new codon encodes a different amino acid, leading to production of a variant haemoglobin. The site of a mutation is critical, determining whether there is an effect on stability, oxygen affinity, solubility and other critical characteristics of the haemoglobin molecule. Because of the redundancy in the genetic code, different point mutations can give rise to the same variant haemoglobin. For example, the α chain variant, G-Philadelphia, has arisen twice, from an AAC to AAG change in an α2α1 fusion gene and from an AAC to AAA change in an α2 gene [88]. There are more than 1000 known variant haemoglobins resulting from point mutations in the α or β genes. There are also at least 40 variant haemoglobins resulting from two point mutations in the same gene with two amino acid substitutions. This can result either from a new mutation occurring in the gene encoding a variant globin chain (e.g. in a parental germ cell) or from crossover between two variant alleles. Usually the second mutation has occurred in a gene that would otherwise encode a relatively common variant haemoglobin such as haemoglobin S, C, E or D-Punjab. Strangely, double mutations have been much more frequently described in the β gene than in the α genes.

Some point mutations are 'nonsense' mutations in which the new codon is one of the three that do not code for an amino acid. A 'nonsense' mutation thus functions as a 'STOP' or 'TERMINATION' codon, leading to termination of chain synthesis. If this type of mutation occurs towards the 5′ end of the gene, no functional protein is produced. However, when the mutation occurs near the 3′ end, an abnormal globin gene may be produced which interferes with the formation of functional haemoglobin tetramers and leads to a dominant thalassaemia phenotype. Point mutations can also convert a STOP codon to a coding sequence so that an elongated mRNA and elongated globin chain are produced.

An unusual result of a point mutation is production of an abnormal amino acid that is converted to a different amino acid by post-translational modification. This can be as the result of deamidation, acetylation or oxidation. There are at least six reported variant haemoglobins in which the abnormal DNA sequence codes for asparagine but this is subsequently deamidated to aspartic acid [19]; of these, the most common is haemoglobin J Sardegna ($\alpha 50[CD8]^{His \rightarrow Asn \rightarrow Asp}$), which has a prevalence of 0.25% in northern Sardinia. Post-translational acetylation occurs in haemoglobin Raleigh, which has a $\beta 1^{Val \rightarrow Ala}$ substitution; proteins with an N-terminal alanine are often acetylated and this is the case with this variant haemoglobin [89]. The presence in one individual of haemoglobins with three different β chains may be attributable to post-translational modification. For example, the replacement of leucine by hydroxyleucine that characterises haemoglobin Coventry is not encoded by genomic DNA and is found only in the presence of an unstable haemoglobin, either haemoglobin Atlanta or haemoglobin Sydney. Some mutations affecting the haem pocket and leading to haemoglobin instability permit oxidation of leucine to isoleucine [90]. Haemoglobin Bristol also shows post-translational modification. It is an unstable haemoglobin resulting from conversion of the β67 valine codon to a codon for methionine; however, the final haemoglobin has aspartic acid rather than methionine as a result of post-translational modification [19].

In a slightly different mechanism, the abnormal structure of a variant haemoglobin resulting from a point mutation leads to post-translational modification of a normal amino acid, in three cases leucine being modified to hydroxyleucine [19] and in one case asparagine

adjacent to the abnormal residue being deamidated to aspartic acid [89].

Mutations in the codon for the N-terminal valine may mean that a different amino acid is encoded with resultant retention of the initiator methionine and full acetylation of the N-terminal residue (e.g. the glutamate of the α chain variant haemoglobin Thionville) or normal cleavage of methionine but full acetylation of the N-terminal residue (e.g. the alanine of the α chain variant haemoglobin Lyon-Bron) [40]. Similarly, a histidine to proline change in position β2 leads to retention of the initiator methionine [89]. If methionine is retained, the globin chain is extended by one residue.

Deletions and insertions can lead to a frame shift; unless the deletion or insertion involves three nucleotides or multiples of three the nucleotide sequences beyond the mutation will be in a different reading frame and will be 'read' during translation as coding for a completely different sequence of amino acids. Frame shift mutations can lead to a premature STOP codon so that both mRNA and the resultant globin chain are shorter than normal. Unless this occurs, a frame shift mutation is likely to lead to elongated mRNA and an elongated globin chain. The original STOP codon is no longer in the reading frame and transcription continues until another STOP codon is encountered.

Small deletions and large deletions and insertions can result from non-homologous crossover between a pair of chromosomes during meiosis. These are usually in-frame. Non-homologous crossover can involve not only a single pair of allelic genes (e.g. two α genes) but also two structurally similar but non-allelic genes (e.g. a β gene and a δ gene); in the latter instance there may be a loss of the two normal genes and production of a fusion gene that has 5′ sequences of one gene and 3′ sequences of the other gene; alternatively the two normal genes may be retained with part of both genes being reduplicated in the fusion gene. Some examples of non-homologous crossover are shown in Fig. 1.14. Non-homologous crossover can also result in

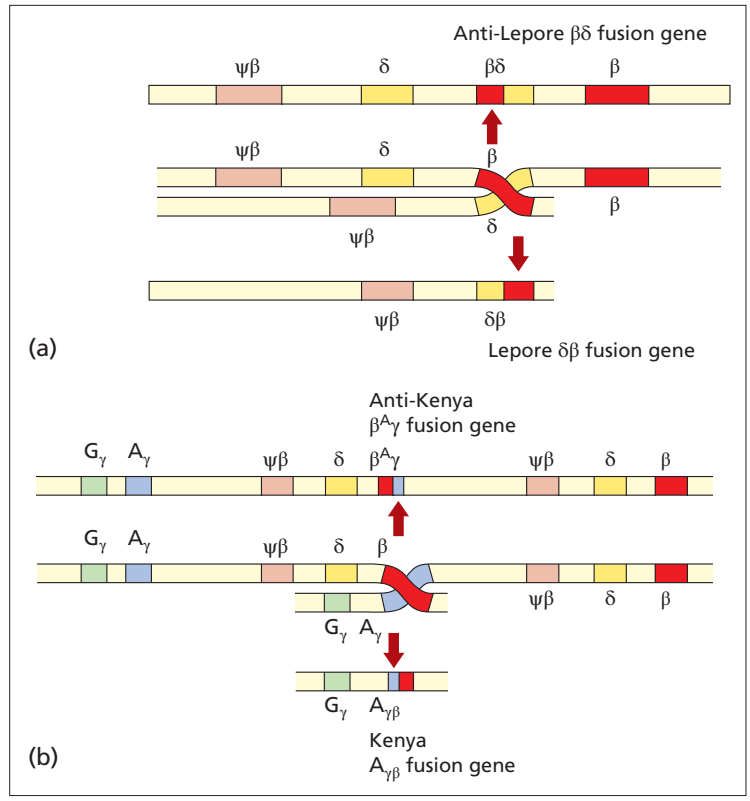

Fig. 1.14 Some examples of fusion genes produced by non-homologous crossover: (a) formation of genes encoding Lepore and anti-Lepore haemoglobins; (b) formation of genes encoding Kenya and anti-Kenya haemoglobins.

the reduplication or loss of some genes, referred to as copy number variants (CNVs); for example, some individuals, instead of having two α genes on each chromosome 16, have three, four or even five [91] α genes on one chromosome. Duplicated α genes occur in many populations and in some are quite frequent. For example, 2% of Sri Lankans have ααα.

Very rarely, individuals are somatic mosaics so that a variant haemoglobin is present in an unusually low percentage. For example, haemoglobin Costa Rica, a β chain variant arising as a result of somatic mutation, constituted 5–6% of total haemoglobin in the individual in whom it was recognised [92]. Similarly, a patient has been reported with haemoglobin Korle Bu as a minor fraction as a result of constitutional mosaicism [93].

Haemoglobin dimers are stable but the tetramers that they form are able to dissociate and re-associate. When both normal and variant globin chains are present, heterotetramers and homotetramers will be present *in vivo* (e.g. in the case of sickle cell trait there will be $\alpha_2\beta_2$, $\alpha_2\beta^S{}_2$ and $\alpha_2\beta\beta^S$). When haemoglobins are studied *in vitro* (e.g. by electrophoresis or chromatography), the heterotetramers dissociate and re-associate as homotetramers. Some variant haemoglobins have abnormally stable tetramers so that three rather than two forms are detected on haemoglobin electrophoresis and similar techniques.

Thalassaemias and variant haemoglobins

Mutations can lead not only to synthesis of a structurally abnormal haemoglobin but also to a reduced rate of synthesis of a globin chain and therefore of the haemoglobin species of which it forms a part. The term 'thalassaemia' is used to describe disorders with a significant decrease in the rate of synthesis of one or more globin chains. α thalassaemia indicates a reduced rate of synthesis of α globin chain. Similarly, β, δ, δβ and γδβ thalassaemias indicate a reduced rate of synthesis of β, δ, δ + β and γ + δ + β chains respectively. In some disorders there is both synthesis of a structurally abnormal haemoglobin and a reduced rate of synthesis of the variant haemoglobin. This is the case, for different reasons, with the α chain variant, haemoglobin Constant Spring (first described in a Chinese patient in Constant Spring, a district of Kingston, Jamaica), and the β chain variant, haemoglobin E. The term 'haemoglobinopathy' is sometimes used to indicate only those disorders with a structurally abnormal haemoglobin, whereas others use the term to include all disorders of globin chain synthesis, encompassing also the thalassaemias. The second use seems preferable since some variant haemoglobins, such as haemoglobin E and haemoglobin Constant Spring, are synthesised at a reduced rate and thus constitute thalassaemic haemoglobinopathies, and some thalassaemias lead to synthesis of a structurally abnormal haemoglobin, such as haemoglobin H or haemoglobin Bart's, as a result of unbalanced chain synthesis.

Haemoglobinopathies can result from mutation of a β globin gene, in which case there is only a variant form of haemoglobin A, or mutation of an α globin gene, in which case there are variant forms of haemoglobins F, A and A_2. Similarly, mutations of γ or δ genes result in mutant forms of haemoglobin F and haemoglobin A_2 respectively. Because there are two β genes, an individual can have two different β globin variants. Because there are usually four α genes, an individual could, in theory, have up to four different α chain variants; in practice, a number of individuals have been described with many different haemoglobin variants, for example, haemoglobin Buda, haemoglobin Pest and haemoglobin A in one instance and haemoglobin $G^{Philadelphia}$, haemoglobin $J^{Sardegna}$ and haemoglobin A in several instances. These tend to be more common in countries, such as Thailand, with a high incidence of many different variants.

The proportion of variant haemoglobins

The proportion of an α chain variant in the blood might be expected to be around 25%, since there are usually four α genes. However, the situation is far more complex. The variant is likely to be more than 25% if it results from a mutation of the α2 gene (since the ratio of α2 to α1 synthesis is normally about 3:1) and less

Table 1.5 Consequences of mutations of globin genes.

Type of mutation and consequence	Examples
Substitution of an external amino acid which is not involved in inter-chain contacts; no functional abnormality	Haemoglobin G-Philadelphia
Amino acid substitution leading to reduced solubility, polymerisation of haemoglobin and deformation of cells into a holly leaf or sickle shape with consequent haemolysis and vascular obstruction	Haemoglobin S (sickle cell haemoglobin)
Amino acid substitution leading to reduced solubility, formation of straight-edged crystals and haemolysis	Haemoglobin C
Replacement of haem-binding or haem-related histidine residue by another amino acid leading to an increased tendency to oxidation, i.e. formation of methaemoglobin. There is cyanosis at birth if the defect is in a γ gene, cyanosis from birth if the defect is in an α chain and cyanosis from approximately 6 months of age if the defect is in a β chain. There may be associated haemoglobin instability	M haemoglobins
Mutation involving amino acids of the haem pocket or $\alpha_1\beta_2$ (tetrameric) contacts or mutation interfering with the helical structure of haemoglobin, leading to haemoglobin instability and Heinz body haemolytic anaemia; there may also be decreased oxygen affinity and resultant cyanosis	Haemoglobin Köln, haemoglobin Zurich (haem pocket mutation), haemoglobin Kansas (mutation affecting α1β2 contacts)
Mutations involving $\alpha_1\beta_2$, $\alpha_2\beta_1$ tetrameric haemoglobin contacts or C-terminal end of β chain, where there are residues involved in 2,3-DPG interaction and stability of the deoxy form of haemoglobin, leading to increased oxygen affinity and polycythaemia	Haemoglobin Chesapeake, haemoglobin Bethesda, haemoglobin Kempsey, haemoglobin J-Cape Town, haemoglobin Yakima
Mutation leading to decreased oxygen affinity and therefore anaemia, since normal tissue delivery of oxygen is achieved with a lower concentration of haemoglobin. May cause cyanosis	Haemoglobin S, haemoglobin Seattle (also unstable), haemoglobin Kansas (also unstable), haemoglobin Beth Israel
Mutation in β gene leading to markedly reduced or absent β chain production, reduced synthesis of haemoglobin A and possibly ineffective erythropoiesis consequent on damage to developing erythroblasts by excess α chains	β thalassaemia (major, intermedia or minor)
Mutation in β gene leading to truncated and very unstable β chain	(Dominant) β thalassaemia phenotype
Mutation in one or more α genes leading to markedly reduced or absent α chain synthesis and reduced synthesis of haemoglobins F, A and A_2	α thalassaemia (α thalassaemia trait, haemoglobin H disease or haemoglobin Bart's hydrops fetalis)
Mutation in α gene leading to structurally abnormal α chain synthesised at a greatly reduced rate	α thalassaemia phenotype, e.g. haemoglobin Constant Spring (mRNA and the haemoglobin are unstable)
Mutation in δ gene leading to a structural abnormality or markedly reduced or absent δ chain production	Haemoglobin A_2 variant or δ thalassaemia. No clinical significance as haemoglobin A_2 is a minor haemoglobin but complicates the diagnosis of β thalassaemia trait
Mutation in γ gene leading to structural abnormality or reduced rate of synthesis of γ chain and therefore haemoglobin F	Low haemoglobin F levels or haemoglobin F variant

2,3-DPG, 2,3-diphosphoglycerate; mRNA, messenger ribonucleic acid.

than 25% if it results from mutation of an α1 gene. The percentage is raised if there is coinheritance of α thalassaemia and is lowered if there is coinheritance of triple α (ααα). If a gene encoding an α chain variant is a mutated α1 gene in *cis* with deletion of the α2 gene then it can be upregulated, increasing the percentage further. The percentage is reduced if the variant α chain is synthesised at a reduced rate, if it has a lower affinity for β chains than does the normal α chain or if the variant α chain or the variant haemoglobin is unstable.

Similarly, it might be expected that a β chain variant would be about 50% of total haemoglobin in heterozygotes since there are two β genes. As for α chain variants, the situation is much more complex. The percentage may be above 50% in the case of variants with negatively charged β chains, which have a greater affinity than the normal β chain for the positively charged normal α chains (e.g. haemoglobin J Baltimore or J-Iran); if there is coexisting α thalassaemia, leading to a lack of α chains, the percentage of the variant is even higher. The converse is seen with positively charged β chains, such as β^{S}, β^{C}, $\beta^{O\text{-}Arab}$ and $\beta^{D\text{-}Punjab}$, which have a lower affinity than normal β chains for normal α chains. The percentage of the variant is thus somewhat less than 50% and if there is coexisting α thalassaemia the percentage is even lower. The percentage is also reduced considerably below 50% if there is a reduced rate of synthesis of the variant β (or δβ) chain (e.g. β^{E}, $\delta\beta^{Lepore}$), if the β chain is unstable or if the variant haemoglobin is unstable (e.g. haemoglobin Köln).

An alteration in the amino acid sequence of the globin chains (an alteration in the primary structure of haemoglobin) often has no significant effect on the secondary, tertiary and quaternary structure of haemoglobin; this is the case when the substituted amino acid is of similar size to the normal amino acid, has the same charge and the same hydrophobic or hydrophilic properties and does not have a role in the binding of haem or 2,3-DPG nor in interactions between chains. This is the case for the majority of variant haemoglobins, which have no consequences for the health of the individual. In other cases an alteration in the primary structure of haemoglobin affects the secondary, tertiary or quaternary structure of the molecule, sometimes with very profound effects. Some of the effects of mutations in globin genes are shown in Table 1.5.

At least 1870 variants of the globin genes have been identified. Some 690 of them were initially collated in a single volume [94] and this database is now available electronically, in an updated and greatly expanded form (http://globin.cse.psu.edu/).

Check your knowledge

One to five answers may be correct. Answers to almost all questions can be found in this chapter or can be deduced from information given. Answers are given on page 32.

1.1 The haemoglobin molecule
- (a) requires iron for its synthesis
- (b) is composed of three pairs of globin chains
- (c) alters its structure when oxygen is bound
- (d) is assembled in the cytosol
- (e) binds 2,3 diphosphoglycerate

1.2 Haemoglobin F
- (a) is the major haemoglobin present in the fetus
- (b) has a lower oxygen affinity than haemoglobin A
- (c) is absent in normal adults
- (d) percentage shows a non-Gaussian distribution in the population
- (e) is composed of two α chains and two β chains

1.3 The functions of haemoglobin include
- (a) transport of glucose
- (b) transport of CO_2
- (c) transport of O_2
- (d) buffering
- (e) transport of creatinine to the kidney

1.4 The affinity of haemoglobin for oxygen is decreased by
- (a) fever
- (b) metabolic alkalosis
- (c) binding of CO_2

(d) binding of 2,3 diphosphoglycerate
(e) glycosylation

1.5 When blood circulates through the lungs haemoglobin
(a) is oxidised
(b) takes up oxygen
(c) loses CO_2
(d) takes up water
(e) dissociates into haem and globin

1.6 Structurally abnormal haemoglobins may result from
(a) point mutations
(b) gene fusion
(c) frame shift mutations
(d) mutation of STOP codon to a coding sequence
(e) mutation of a coding sequence to a STOP codon

1.7 Abnormal haemoglobins may
(a) have increased oxygen affinity
(b) have decreased oxygen affinity
(c) be prone to crystallise
(d) be unstable
(e) be abnormally prone to oxidation

1.8 Mutations in globin genes
(a) can occur in α, β, $^{G}\gamma$, $^{A}\gamma$ and δ genes
(b) always result in a structural abnormality of haemoglobin
(c) always have harmful effects
(d) can lead to a reduced rate of globin chain synthesis
(e) can convert one gene to another

1.9 Haemoglobin F
(a) is present, in adult life, in a subset of erythrocytes referred to as F cells
(b) is composed of two α chains and two γ chains, encoded by two pairs of structurally similar α genes and two pairs of structurally similar γ genes
(c) has a sigmoid dissociation curve
(d) constitutes a higher proportion of total haemoglobin in premature than in full-term babies
(e) on average is present at a higher level in women than in men

1.10 Cooperativity is essential for
(a) a sigmoid oxygen dissociation curve
(b) the higher oxygen affinity of haemoglobin F in comparison with haemoglobin A
(c) the Bohr effect
(d) the binding of CO_2 to haemoglobin
(e) conversion of haemoglobin to methaemoglobin

1.11 The proportion of a variant haemoglobin is usually
(a) greater in the case of an α chain variant than a β chain variant
(b) greater in the case of an α chain variant if there is coexisting deletion of an α gene
(c) greater if the variant β chain has a higher affinity for normal α chain than does the normal β chain
(d) greater, in the case of haemoglobin S, if there is coexisting α thalassaemia
(e) greater if the variant haemoglobin is unstable

Further reading and resources

Bain BJ, Wild BJ, Stephens AD and Phelan LA. *Variant Haemoglobins: A Guide to Identification*. Oxford, Wiley-Blackwell, 2010.

The Globin Gene Server, hosted by Pennsylvania State University, USA and McMaster University, Canada. http://globin.cse.psu.edu/

Xenophontos M, Minaidou A, Stephanou C, Tamana S, Kleanthous M and Kountouris P (2023) IthaPhen: an interactive database of genotype-phenotype data for hemoglobinopathies. *HemaSphere*, **7**, e992.

References

1 Attia AM, Ibrahim FA, Abd El-Latif NA, Aziz SW, Abdelmottaleb Moussa SA and Elalfy MS (2018) Determination of human hemoglobin derivatives. *Hemoglobin*, **39**, 371–374.
2 Umbreit J (2007) Methemoglobin – it's not just blue: a concise review. *Am J Hematol*, **82**, 134–144.
3 Gross GS and Lane P (1999) Physiological reactions of nitric oxide and hemoglobin: a radical rethink. *Proc Natl Acad Sci USA*, **96**, 9967–9969.

4 Nagel RL and Jaffé ER. CO-, NO-, met-, and sulf-hemoglobinemias: the dyshemoglobinemias. In: Steinberg MH, Forget BG, Higgs DR and Nagel RL eds. *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, 2001, pp. 1214–1233.
5 Schechter AN and Gladwin MT (2003) Hemoglobin and the paracrine and endocrine functions of nitric oxide. *N Engl J Med*, **348**, 1483–1485.
6 Straub AC, Lohman AW, Billaud M, Johnstone SR, Dwyer ST, Lee MY, *et al*. (2012) Endothelial cell expression of haemoglobin α regulates nitric oxide signalling. *Nature*, **491**, 473–477.
7 Denton CC, Shah P, Suriany S, Liu H, Thuptimdang W, Sunwoo J, *et al*. (2021) Loss of alpha-globin genes in human subjects is associated with improved nitric oxide-mediated vascular perfusion. *Am J Hematol*, **96**, 277–281.
8 Pauling L, Itano HA, Singer SY and Wells IG (1949) Sickle-cell anemia, a molecular disease. *Science*, **110**, 543–548.
9 Kunkel HG and Wallenius G (1955) New hemoglobin in normal adult blood. *Science*, **122**, 288.
10 Bunn HF and Forget BG. *Hemoglobin: Molecular, Genetic and Clinical aspects*. Philadelphia, W.B. Saunders, 1986.
11 Huehns ER, Flynn FV, Butler EA and Beaven GH (1961) Two new hemoglobin variants in the very young human embryo. *Nature*, **189**, 496–497.
12 Hecht F, Jones RT and Koler RD (1967) Newborn infants with Hb Portland 1, an indicator of α-chain deficiency. *Ann Hum Genet*, **31**, 215–218.
13 Fantoni A, Farace MG and Gambari R (1981) Embryonic hemoglobins in man and other mammals. *Blood*, **57**, 623–633.
14 Nagel RL and Steinberg MH. Hemoglobins of the embryo and fetus and minor hemoglobins of adults. In: Steinberg MH, Forget BG, Higgs DR and Nagel RL eds. *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, 2001, pp. 197–230.
15 Kabadi UM, Gopal V, Hood L, Kabadi MU and Platt K (1992) Elevated glycosylated hemoglobin concentrations in AIDS. *AIDS*, **6**, 236–238.
16 Zurbriggen K, Schmugge M, Schmid M, Durka S, Kleinert P, Kuster T, *et al*. (2002) Analysis of minor hemoglobins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem*, **51**, 989–996.
17 Naito C and Niwa T (2000) Analysis of glutathionyl hemoglobin levels in diabetic patients by electrospray ionization liquid chromatography-mass spectrometry: effect of vitamin E administration. *J Chromatogr B Biomed Sci Appl*, **746**, 91–94.
18 Niketić V, Beslo D, Rajcević S, Sredić S and Stojković M (1992) Glutathione adduct of hemoglobin (Hb ASSG) in hemolysates of patients on long-term antiepileptic therapy. *Int J Biochem*, **24**, 503–507.
19 Rees DC, Rochette J, Schofield C, Green B, Morris M, Parker NE, *et al*. (1996) A novel silent posttranslational mechanism converts methionine to aspartate in hemoglobin Bristol (β67[E11] Val-Met→Asp). *Blood*, **88**, 341–348.
20 Mitchell TR and Pegrum GD (1971) The oxygen affinity of haemoglobin in chronic renal failure. *Br J Haematol*, **21**, 463–472.
21 Ponka P (1997) Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood*, **89**, 1–25.
22 Chen J-J (2007) Regulation of protein synthesis by the heme-regulated eIF2α kinase: relevance to anemias. *Blood*, **109**, 2693–2699.
23 Steinberg MH and Adams JG (1991) Haemoglobin A_2: origin, evolution, and aftermath. *Blood*, **78**, 2165–2177.
24 Roberts AV, Weatherall DJ and Clegg JB (1972) The synthesis of human haemoglobin A_2 during erythroid maturation. *Biochem Biophys Res Commun*, **47**, 81–87.
25 Stamatoyannopoulos G and Nienhuis AW. Hemoglobin switching. In: Stamatoyannopoulos G, Nienhuis AW, Majerus PW and Varmus H eds. *The Molecular Basis of Blood Diseases*, 2nd edn. W.B. Saunders, Philadelphia, 1994, pp. 107–156.
26 Weatherall D and Clegg JB. *The Thalassaemia Syndromes*, 4th edn. Oxford, Blackwell Science, 2001.
27 Badat M, Davies JOJ, Fisher CA, Downes DJ, Rose A, Glenthøj AB, *et al*. (2021) A remarkable case of HbH disease illustrates the relative contributions of the α-globin enhancers to gene expression. *Blood*, **137**, 572–575.

28 Ho PJ and Thein SL (2000) Gene regulation and deregulation: a β globin perspective. *Blood Rev*, **14**, 78–93.

29 Blayney JW, Francis H, Rampasekova A, Camellato B, Mitchell L, Stolper R, *et al*. (2023) Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell*, **86**, 5826–5839.

30 Russell JE and Liebhaber SA (1996) The stability of human β-globin mRNA is dependent on structural determinants positioned within its 3′ untranslated region. *Blood*, **87**, 5314–5322.

31 Jane SM and Cunningham JM (1998) Understanding fetal globin gene expression: a step towards effective Hb F reactivation in haemoglobinopathies. *Br J Haematol*, **102**, 415–422.

32 Sankaran VG, Xu J, Byron R, Greisman HA, Fisher C, Weatherall DJ, *et al*. (2011) A functional element necessary for fetal hemoglobin silencing. *N Engl J Med*, **365**, 807–814.

33 Liu D, Zhang X, Yu L, Cai R, Ma X, Zheng C, *et al*. (2014) KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of β-thalassemia. *Blood*, **124**, 803–811.

34 Funnell AP, Prontera P, Ottaviani V, Piccione M, Giambona A, Maggio A, *et al*. (2015) 2p15-p16.1 microdeletions encompassing and proximal to *BCL11A* are associated with elevated HbF in addition to neurologic impairment. *Blood*, **126**, 89–93.

35 Jiang F, Chen GL, Li J, Zhou JY, Liao C and Li DZ (2018) Analysis of the genotypes in a Chinese population with increased Hb A_2 and low hematological indices. *Hemoglobin*, **42**, 161–165.

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

37 Basu P, Morris PE, Haar JL, Wani MA, Lingrel JB, Gaensler KM and Lloyd JA (2005) KLF2 is essential for primitive erythropoiesis and regulates the human and murine embryonic beta-like globin genes in vivo. *Blood*, **106**, 2566–2571.

38 Achour A, Koopmann T, Castel R, Santen GWE, den Hollander N, Knijnenburg J, *et al*. (2020) A new gene associated with a β-thalassemia phenotype: the observation of variants in *SUPT5H*. *Blood*, **136**, 1789–1793.

39 Boissel JP, Kasper TJ, Shah SC, Malone JI and Bunn HF (1985) Amino-terminal processing of proteins: hemoglobin South Florida, a variant with retention of initiator methionine and N alpha-acetylation. *Proc Natl Acad Sci USA*, **82**, 8448–8452.

40 Lacan P, Souillet G, Aubry M, Prome D, Richelme-David S, Kister J, *et al*. (2002) New alpha 2 globin chain variant with low oxygen affinity affecting the N-terminal residue and leading to N-acetylation [Hb Lyon-Bron α 1(NA1)Val --> Ac-Ala]. *Am J Hematol*, **69**, 214–218.

41 Bernini LF and Harteveld CL (1998) α-thalassaemia. *Bailliere's Clin Haematol*, **11**, 53–90.

42 Serjeant BE, Mason KP and Serjeant GR (1978) The development of haemoglobin A_2 in normal Negro infants and in sickle cell disease. *Br J Haematol*, **39**, 259–263.

43 Steinberg MH (1998) Pathophysiology of sickle cell disease. *Bailliere's Clin Haematol*, **11**, 163–184.

44 Steinberg MH and Nagel RL. Hemoglobins of the embryo, fetus, and adult. In: Steinberg MH, Forget BG, Higgs DR and Weatherall DJ eds. *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*, 2nd edn. Cambridge University Press, Cambridge, 2009, pp. 119–135.

45 Menzel S, Garner C, Rooks H, Spector TD and Thein SL (2013) HbA_2 levels in normal adults are influenced by two distinct genetic mechanisms. *Br J Haematol*, **160**, 101–105.

46 Sankaran VG, Joshi M, Agrawal A, Schmitz-Abe K, Towne MC, Marinakis N, *et al*. (2013) Rare complete loss of function provides insight into a pleiotropic genome-wide association study locus. *Blood*, **122**, 3845–3847.

47 Perseu L, Satta S, Moi P, Demartis FR, Manunza L, Sollaino MC, *et al*. (2011) *KLF1* gene mutations cause borderline Hb A2. *Blood*, **118**, 4454–4458.

48 Nitta T, Kawano F, Yamashiro Y, Takagi F, Murata T, Tanaka T, *et al*. (2015) A new Krüppel-like factor 1 mutation (c.947G>A or p.C316Y) in humans causes β-thalassemia minor. *Hemoglobin*, **39**, 121–126.

49 Borbely N, Phelan L, Szydlo R and Bain B (2013) Capillary zone electrophoresis for haemoglobinopathy diagnosis. *J Clin Pathol*, **66**, 29–39.

50 Kaeda JS, Prasad K, Howard RJ, Mehta A, Vulliamy T and Luzzatto L (1994) Management

of pregnancy when maternal blood has a very high level of fetal haemoglobin. *Br J Haematol*, **88**, 432–434.

51 Huehns ER and Beaven GH. Developmental changes in human haemoglobins. In: Benson PF ed. *The Biochemistry of Human Development*. Spastics International Medical, London, 1971, p. 175.

52 Colombo B, Kim B, Atencio RP, Molina C and Terrenato L (1976) The pattern of foetal haemoglobin disappearance after birth. *Br J Haematol*, **32**, 79–87.

53 Cheron G, Bachoux I, Maier M, Massonneau M, Peltier JY and Girot T (1989) Fetal hemoglobin in sudden infant death syndrome. *N Engl J Med*, **320**, 1011–1012.

54 Garner C, Tatu T, Reittie JE, Littlewood T, Darley J, Cervino S, *et al*. (2000) Genetic influences on F cells and other hematologic variables: a twin heritability study. *Blood*, **95**, 342–346.

55 Miyoshi K, Kaneto Y, Kawai H, Ohchi H, Niki S, Hasegawa K, *et al*. (1988) X-linked dominant control of F-cells in normal adult life: characterization of the Swiss type as hereditary persistence of fetal hemoglobin regulated dominantly by gene(s) on X chromosome. *Blood*, **72**, 1854–1860.

56 Merghoub T, Perichon B, Maier-Redelsperger M, Dibenedetto SP, Samperi P, Ducrocq R, *et al*. (1997) Dissection of the association status of two polymorphisms in the β-globin gene cluster with variations in F-cell number in non-anemic individuals. *Am J Hematol*, **56**, 239–243.

57 Craig JE, Rochette J, Sampietro M, Wilkie AOM, Barnetson R, Hatton CSR, *et al*. (1997) Genetic heterogeneity in heterocellular hereditary persistence of fetal hemoglobin. *Blood*, **90**, 428–434.

58 Garner C, Silver N, Best S, Menzel S, Martin C, Spector TD and Thein SL (2004) Quantitative trait locus on chromosome 8q influences the switch from fetal to adult hemoglobin. *Blood*, **104**, 2184–2186.

59 Akinsheye I, Alsultan A, Solovieff N, Ngo D, Baldwin CT, Sebastiani P, *et al*. (2011) Fetal hemoglobin in sickle cell anemia. *Blood*, **118**, 19–27.

60 Pereira C, Relvas L, Bento C, Abade A, Ribeiro ML and Manco L (2015) Polymorphic variations influencing fetal hemoglobin levels: association study in β-thalassemia carriers and in normal individuals of Portuguese origin. *Blood Cells Mol Dis*, **54**, 315–320.

61 Kesornsit A, Jeenduang N, Horpet D, Plyduang T and Nuinoon M (2018) Quantitative trait loci influencing Hb F levels in Southern Thai Hb E (*HBB*: c.79G>A) heterozygotes. *Hemoglobin*, **42**, 23–29.

62 Pakdee N, Yamsri S, Fucharoen G, Sanchaisuriya K, Pissard S and Fucharoen S (2014) Variability of hemoglobin F expression in hemoglobin EE disease: hematological and molecular analysis. *Blood Cells Mol Dis*, **53**, 11–15.

63 Alter BP, Rosenberg PS, Day T, Menzel S, Giri N, Savage SA and Thein SL (2013) Genetic regulation of fetal haemoglobin in inherited bone marrow failure syndromes. *Br J Haematol*, **162**, 542–546.

64 Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, *et al*. (2014) HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. *J Clin Invest*, **124**, 1699–1710.

65 Urio F, Nkya S, Rooks H, Mgaya JA, Masamu U, Sangeda RZ, *et al*. (2020) F cell numbers are associated with an X-linked genetic polymorphism and correlate with haematological parameters in patients with sickle cell disease. *Br J Haematol*, **191**, 888–896.

66 NickAria S, Haghpanah S, Ramzi M and Karimi M (2018) Relationship of the interaction between two quantitative trait loci with γ-globin expression in β-thalassemia intermedia patients. *Hemoglobin*, **42**, 108–112.

67 Singleton BK, Lau W, Fairweather VS, Burton NM, Wilson MC, Parsons SF, *et al*. (2011) Mutations in the second zinc finger of human EKLF reduce promoter affinity but give rise to benign and disease phenotypes. *Blood*, **118**, 3137–3145.

68 Satta S, Perseu L, Moi P, Asunis I, Cabriolu A, Maccioni L, *et al*. (2011) Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica*, **96**, 767–770.

69 Petruzzelli R, Gaudino S, Amendola G, Sessa R, Puzone S, Di Concilio R, *et al*. (2010) Role of the cold shock domain protein A in the transcriptional regulation of HBG expression. *Br J Haematol*, **150**, 689–699.

70 Habara AH, Shaikho EM and Steinberg MH (2017) Fetal hemoglobin in sickle cell anemia: the Arab-Indian haplotype and new therapeutic agents. *Am J Hematol*, **92**, 1233–1242.
71 Huang P, Keller CA, Giardine B, Grevet JD, Davies JOJ, Hughes JR, *et al*. (2017) Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev*, **31**, 1704–1713.
72 Ivaldi MS, Diaz LF, Chakalova L, Lee J, Krivega I and Dean A (2018) Fetal γ-globin genes are regulated by the *BGLT3* long noncoding RNA locus. *Blood*, **132**, 1963–1973.
73 Al-Ali ZA, Fallatah RK, Aljaffer EA, Albukhari ER, Sadek Al-Ali N, *et al*. (2018) *ANTXR1* intronic variants are associated with fetal hemoglobin in the Arab-Indian haplotype of sickle cell disease. *Acta Haematol*, **140**, 55–59.
74 Gong Y, Zhang X, Zhang Q, Zhang Y, Ye Y, Yu W, *et al*. (2021) A natural *DNMT1* mutation elevates the fetal hemoglobin level via epigenetic derepression of the gamma-globin gene in beta-thalassemia. *Blood*, **137**, 1652–1657.
75 Qin K, Lan X, Huang P, Saari MS, Khandros E, Keller CA, *et al*. (2023) Molecular basis of polycomb group protein-mediated fetal hemoglobin repression. *Blood*, **14**, 2756–2770.
76 Tadmouri GO, Yüksel L and Basak AN (1998) Hb S/beta (del) thalassemia associated with high levels of hemoglobin A_2 and F in a Turkish family. *Am J Hematol*, **59**, 83–86.
77 Ibrahim M, Qari MH, Sait W and Abulela M (2009) Pattern of HB F level rise during normal pregnancies. *Hemoglobin*, **33**, 534–538.
78 Yamada T, Morikawa M, Yamada T, Nishida R, Takeda M, Kawaguchi S and Minakami H (2013) Changes in hemoglobin F levels in pregnant women unaffected by clinical fetomaternal hemorrhage. *Clin Chim Act*, **415**, 124–127.
79 Zertal-Zidani S, Ducrocq R, Weil-Oliver C, Elion J and Krishnamoorthy R (2001) A novel δβ fusion gene expresses hemoglobin A (HbA) not Hb Lepore: Senegalese $\delta^0\beta^+$ thalassaemia. *Blood*, **98**, 1261–1263.
80 Kim SY, Lee SH, Cho SI, Song SH, Hattori Y, Park S-K, *et al*. (2010) Molecular identification of the novel Gγ-β hybrid hemoglobin: Hb Gγ-β Ulsan (Gγ through 13; β from 19). *Int J Lab Hematol*, **45**, 276–279.
81 Blacklock HA, Case J, Chan T, Raizis T, Doocey R, Fellowes A, *et al*. (2005) Novel sequence insertion in a Maori patient with transfusion-dependent beta-thalassaemia. *Br J Haematol*, **131**, 400–402.
82 De Gobbi M, Viprakasit V, de Jong PJ, Yoshinaga Y, Cheng J-F, Hughes JR, *et al*. (2005) Identification of a gain-of-function SNP causing a new model of α-thalassemia. *Blood*, **106**, 156a.
83 Viprakasit V, Gibbons RJ, Broughton BC, Tolmie JL, Brown D, Lunt P, *et al*. (2001) Mutations in the general transcription factor TFIIH result in beta-thalassaemia in individuals with trichothiodystrophy. *Hum Mol Genet*, **10**, 2797–2802.
84 Yu C, Niakan KK, Matsushita M, Stamatoyannopoulos G, Orkin SH and Raskind WH (2002) X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood*, **100**, 2040–2045.
85 Phillips JD, Steensma DP, Spangrude GJ and Kushner JP (2005) Congenital erythropoietic porphyria, β-thalassemia intermedia and thrombocytopenia due to a GATA1 mutation. *Blood*, **106**, 154a.
86 Breton A, Theodorou A, Aktuna S, Sonzogni L, Darling D, Chan L, *et al*. (2016) ASH1L (a histone methyltransferase protein) is a novel candidate globin gene regulator revealed by genetic study of an English family with beta-thalassaemia unlinked to the beta-globin locus. *Br J Haematol*, **175**, 525–530.
87 Faà V, Meloni A, Moi L, Ibba G, Travi M, Vitucci A, *et al*. (2006) Thalassaemia-like carriers not linked to the β-globin gene cluster. *Br J Haematol*, **132**, 640–650.
88 Huisman THJ (1997) Gamma chain abnormal human fetal hemoglobin variants. *Am J Hematol*, **55**, 159–163.
89 Steinberg MH and Nagel RL. New and recombinant mutant hemoglobins of biological interest. In: Steinberg MH, Forget BG, Higgs DR and Nagel RL eds. *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. Cambridge University Press, Cambridge, 2001, pp. 1195–1211.
90 Brennan SO, Shaw J, Allen J and George PM (1992) β141 Leu is not deleted in the unstable haemoglobin Atlanta-Coventry but is replaced by a novel amino acid of mass 128 daltons. *Br J Haematol*, **81**, 99–103.

91 Cook RJ, Hoyer JD and Highsmith WE (2006) Quintuple α-globin gene: a novel allele in a Sudanese man. *Hemoglobin*, **30**, 51–55.

92 Rodriguez Romero WE, Castillo M, Chaves MA, Saenz GF, Gu LH, Wilson JB, *et al*. (1996) Hb Costa Rica or alpha 2 beta 2 77(EF1)His --> Arg: the first example of a somatic cell mutation in a globin gene. *Hum Genet*, **97**, 829–833.

93 Wild BJ, Green BN, Lalloz MRA and Layton DM (2000) When is a minor haemoglobin fraction worthy of investigation? *Br J Haematol*, **108**, Suppl. 1, 40.

94 Huisman THJ, Carver MFH and Efremov GD. *A Syllabus of Human Hemoglobin Variants*. Augusta, GA, Sickle Cell Anemia Foundation, 1996.

Answers to questions

1.1	(a) T	1.3	(a) F	1.5	(a) F	1.7	(a) T	1.9	(a) T	1.11	(a) F
	(b) F		(b) T		(b) T		(b) T		(b) T		(b) T
	(c) T		(c) T		(c) T		(c) T		(c) T		(c) T
	(d) T		(d) T		(d) F		(d) T		(d) T		(d) F
	(e) T		(e) F		(e) F		(e) T		(e) T		(e) F
1.2	(a) T	1.4	(a) T	1.6	(a) T	1.8	(a) T	1.10	(a) T		
	(b) F		(b) F		(b) T		(b) F		(b) F		
	(c) F		(c) T		(c) T		(c) F		(c) T		
	(d) T		(d) T		(d) T		(d) T		(d) F		
	(e) F		(e) F		(e) T		(e) T		(e) F		