ONE

GLOBIN GENES AND HAEMOGLOBIN

Normal haemoglobins and their synthesis

Haemoglobin is the major protein in the red blood cell. It is a transport protein for oxygen and thus is essential for life. Not all haemoglobin in the human body is the same. During adult life, the major haemoglobin, known as haemoglobin A, comprises about 97% of total haemoglobin. Minor components are haemoglobin A_2 and haemoglobin F. During embryonic and fetal life the situation is very different. The embryo has mainly haemoglobins Gower 1, Gower 2 and Portland 1 whereas fetal life is characterized by synthesis of haemoglobin F and increasingly, as gestation proceeds, haemoglobin A.

All normal haemoglobins are composed of two unlike pairs of polypeptide chains known as globin chains, each of which provides a pocket for an iron-containing haem molecule; the globin protects haem from oxidation. It is the different globin chain composition and the interaction between chains that gives the various haemoglobins their differing characteristics. The normal haemoglobins and their constituent chains are summarized in Table 1.1.

Globin chains are encoded by globin genes, which are located in two clusters, one on chromosome 16 and the other on chromosome 11. The α globin cluster is located near the telomere of chromosome 16 and includes a ζ gene and two α genes, in addition to a number of pseudogenes. There is an upstream positive regulatory region designated the locus control region, alpha (*LCRA*) or HS −40 (since the region is hypersensitive to DNase and is 40 kb upstream of the α globin cluster). The β cluster is located on chromosome 11 and includes an ε gene, two γ genes, a δ gene and a β gene. It also has an upstream positive regulatory region designated the locus control region, beta (*LCRB*). These two gene clusters are shown diagrammatically in Figure 1.1.

The synthesis of haemoglobin is complex. Haem is synthesized partly within mitochondria and partly in the cytosol, a total of eight enzymes being required. Its basic structure is that of a porphyrin ring with a Fe^{++} (ferrous iron) atom at its centre. Globin chains, like all polypeptides, are synthesized on ribosomes, with α chains being synthesized somewhat in excess of β chains. An α chain is thus able to combine with a β chain that is still attached to its ribosome, to form a dimer, which is then detached. Each globin chain of the dimer incorporates a haem molecule before the dimer associates with another dimer to form a haemoglobin tetramer. The tetrameric structure of haemoglobin is fundamental for its function.

Haemoglobin has a primary structure (the sequence of amino acids), a secondary structure (the alternation of α helixes and non-helical turns), a tertiary structure (the three-dimensional arrangement of the haemoglobin monomer) and a quaternary structure (the relationship of the four haemoglobin monomers to each other in the tetramer). An alteration in the primary structure can affect the secondary, tertiary and quaternary structure of haemoglobin. The tetrameric structure (Figure 1.2) is a major evolutionary improvement on more primitive oxygen-binding proteins. The ability of the monomers to alter their relationship to each other on oxygen binding or dissociation is known as co-operativity. Its effect is that the uptake of oxygen by one monomer facilitates uptake by other monomers, and similarly, release of one oxygen facilitates release of the others. The functional importance of this is that in the oxygen-rich environment of the lungs, oxygen is readily

Variant Haemoglobins: a Guide to Identification, 1st edition. By Barbara J. Bain, Barbara J. Wild, Adrian D. Stephens and Lorraine A. Phelan. Published 2010 by Blackwell Publishing Ltd.

Table 1.1 The normal haemoglobins of man.

Haemoglobin	Globin chains	Period of life when present
Gower 1	$\zeta_2\varepsilon_2$	Embryo
Gower 2	$\alpha_2\varepsilon_2$	Embryo
Portland 1	$\zeta_2\gamma_2$	Embryo
Haemoglobin F	$\alpha_2\gamma_2$	Embryo, fetus and neonate; minor component during adult life
Haemoglobin A	$\alpha_2\beta_2$	Minor component in fetus, increasing late in gestation and in the neonatal period to become the major haemoglobin during infancy, childhood and adult life
Haemoglobin A_2	$\alpha_2\delta_2$	Very low levels in infancy; minor component in childhood and adult life

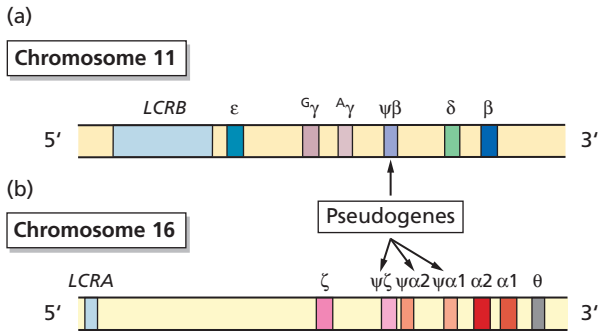

Figure 1.1 Diagram of the α and β globin gene clusters: (a) the β globin gene cluster at 11p15.5 showing the locus control region, beta (*LCRB*), the ε, $^{G}\gamma$, $^{A}\gamma$, δ and β genes and the ψβ pseudogene; (b) the α gene cluster at 16pter-p13.3 showing the locus control region, alpha (*LCRA*), the ζ, α2 and α1 genes and the pseudogenes, ψζ, ψα2 and ψα1, and the θ gene (of unknown functionality).

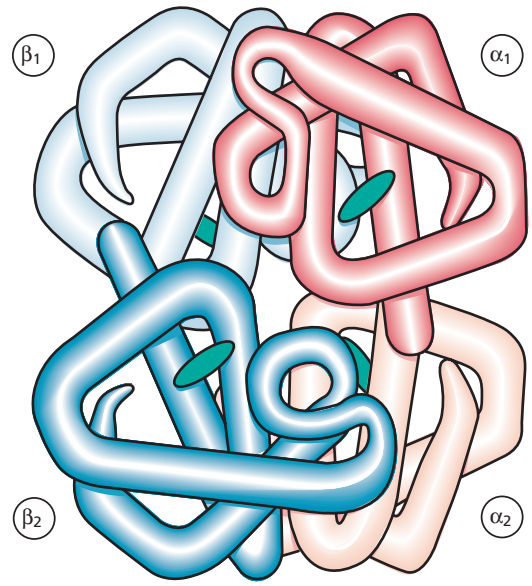

Figure 1.2 Diagram showing the tetrameric structure of haemoglobin A: the $\alpha_1\beta_1$ dimer is at the top and the $\alpha_2\beta_2$ dimer at the bottom; the haem molecules are represented in green.

taken up whereas in conditions of relative hypoxia, in peripheral tissues, oxygen is readily given up. It is this co-operativity that is responsible for the normal sigmoid oxygen dissociation curve of haemoglobin (Figure 1.3). Certain abnormal haemoglobins resemble primitive oxygen-binding proteins in that, in hypoxic conditions, they release oxygen less readily than haemoglobin A and the haemoglobin concentration rises to compensate for this; if co-operativity is entirely lost, the haemoglobin oxygen dissociation curve is hyperbolic.

Although oxygen transport is the major function of haemoglobin it is not the sole function. Haemoglobin also transports CO_2 from tissues to lungs and has a buffering capacity, reducing the swings in pH that could otherwise occur. It also has a role in nitric oxide (NO) transport. Haemoglobin can transport nitric oxide to tissues where is causes vasodilation. However, in pathological conditions, binding of NO to haemoglobin is not necessarily beneficial. When there is intravascular haemolysis, as in sickle cell anaemia, free haemoglobin can scavenge nitric oxide leading to

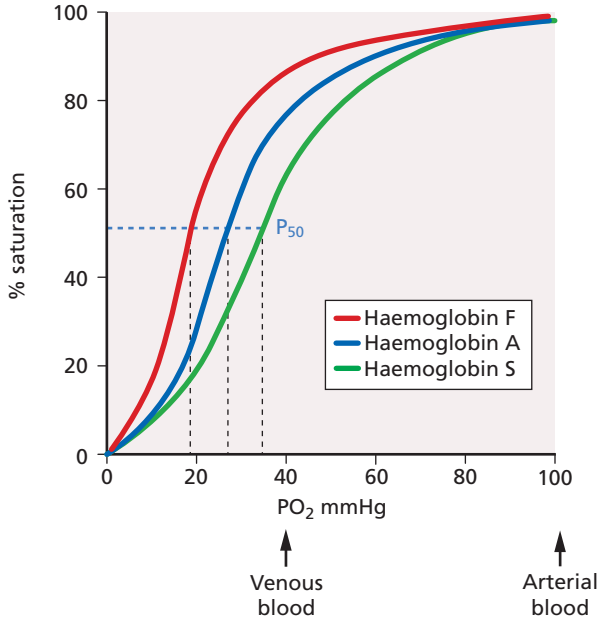

Figure 1.3 Diagram showing the haemoglobin oxygen dissociation curves of haemoglobins A, F and S. Haemoglobin A has a mean P_{50} (partial pressure at which haemoglobin is 50% oxygenated) of about 26.8 mmHg. Haemoglobin S has a lower affinity than haemoglobin A (P_{50} about 35.4 mmHg) whereas haemoglobin F has a higher affinity (P_{50} about 19 mmHg). The partial pressure of oxygen in venous and arterial blood is indicated.

undesirable vasoconstriction, which contributes to pulmonary hypertension.

Globin gene structure and function

In order to understand how a globin gene encodes a globin chain it is necessary to know something of the structure and function of genes. Genes are DNA sequences in which a specific nucleotide sequence carries genetic information. Triplets of nucleotides (codons) either encode specific amino acids or, for a minority of sequences, do not encode an amino acid and thus serve as a stop or termination signal. A functioning gene must commence with a promoter sequence to which transcription factors can bind. This sequence is followed by an initiation sequence, which encodes methionine. Genes are composed of exons, which represent the polypeptide encoded, and introns or intervening sequences, which do not. DNA is present as a double strand, i.e. there are two intertwined strands of DNA with complementary sequences. One of these strands, the 'antisense' strand serves as a template for RNA synthesis so that the messenger RNA (mRNA) that is ultimately produced carries the same genetic message as the 'sense' strand of DNA. In addition to the promoter, which is immediately upstream of the coding sequence of the gene, genes are also influenced by enhancers. These may be located upstream, downstream or even within a gene. In the case of globin genes (and at least three other unrelated genes) there are also upstream sequences that control the transcription of all genes within the cluster, *LCRA* and *LCRB* respectively. In addition, there are various genes encoding transactivating factors, mutation of which is a rare cause of thalassaemia; they include *ATRX* (*XH2*) (α thalassaemia) and *XPD* (also known as *ERCC2*) and *GATA1* (β thalassaemia). There are also two loci, at 6q22.3-23.1 and Xp22.2 respectively, that control the number of haemoglobin F-containing cells (F cells). The genetic control of globin chain synthesis is thus highly complex.

The processes involved in globin chain synthesis are shown diagrammatically in Figure 1.4. The term transcription describes the process by which an RNA precursor molecule is synthesized on a DNA template by means of RNA polymerase. Since both introns and exons are represented in this initial (primary) transcript, further processing is necessary. This processing includes removal of the introns (splicing), addition of an upstream 7-methyl guanosine cap (capping) and addition of a downstream polyadenylate tail (polyadenylation). The 7-methyl guanosine cap appears to have a role during translation. Polyadenylation is important for RNA stability. The result of processing is the production of mRNA. The mRNA moves from the nucleus to the cytoplasm where it serves as a template for ribosomal polypeptide synthesis, a process known as translation. The process also requires transport RNA (tRNA) molecules, which transport the designated amino acid to the growing polypeptide chain on a ribosome. Polypeptide chains normally commence with methionine (represented by ATG in the mRNA), which is subsequently removed. Translation stops when a STOP sequence is encountered in the RNA (TAA, TAG or TGA).

A pseudogene is a DNA sequence, which has occurred during the process of evolution, that resembles a gene in structure but does not lead to the synthesis of a protein. The lack of function

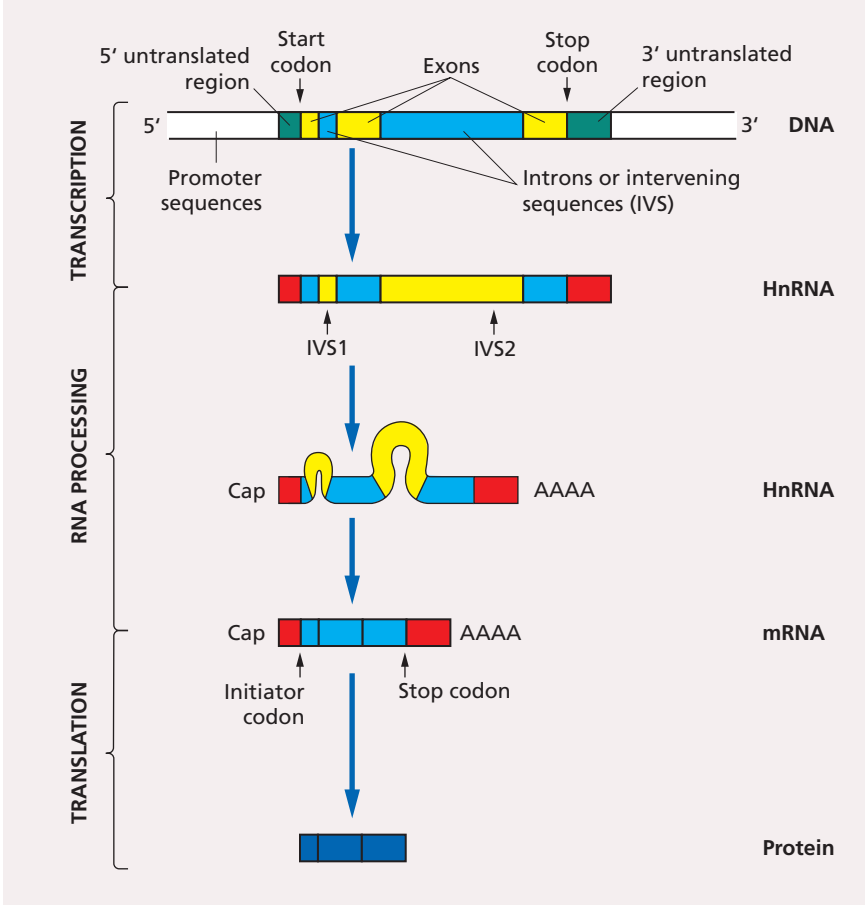

Figure 1.4 Diagram summarizing the processes of transcription, RNA processing and translation. The DNA molecule with a globin gene is represented in line 1. In the process of transcription, a complementary RNA sequence is synthesized on the DNA template. This creates a messenger RNA (mRNA) precursor molecule, known as heterogeneous nuclear RNA (HnRNA), which must be processed by: (i) the addition of a 7-methyl guanosine cap to the 5′ end of the molecule; (ii) splicing out of the introns; and (iii) polyadenylation of the 3′ end of the molecule. Processing leads to formation of mRNA. Processing is followed by translation, in which there is synthesis of a protein on a ribosome, using the mRNA as a template.

may be because of a disabling mutation or because of the lack of a critical element for gene expression. Pseudogenes are transcribed but not translated. Occasionally a further mutation converts a pseudogene into a functioning gene. The globin genes include the gene encoding the δ globin chain, which may be seen as being on its way to becoming a pseudogene; alterations in its promoter have led to a low rate of transcription and consequently haemoglobin A_2 is quite a low proportion of total haemoglobin.

Globin genes are commonly referred to by the same Greek letter as designates the corresponding globin chain. However, they also have 'official' names, as assigned by the Human Genome Project (Table 1.2).

Nomenclature of haemoglobins

Early on, the common haemoglobins found were named as haemoglobin A for adult haemoglobin and haemoglobin F for fetal haemoglobin. Haemoglobin A_2, the minor adult haemoglobin first found on starch block electrophoresis in 1955 [1], was so named in 1957 at a meeting of the International Society of Hematology (ISH) [2]. The same group noted that a minor haemoglobin band was often present slightly anodal to haemoglobin A on starch block electrophoresis at alkaline pH [3]; it was named haemoglobin A_3 at the same ISH meeting [2].

Analysis by cation exchange column chromatography showed that haemoglobin A could be

Table 1.2 The globin genes and locus control genes.

Type of gene	Commonly used name		Official name
Structural genes	Zeta	ζ	*HBZ*
	Alpha 2	α2	*HBA2*
	Alpha 1	α1	*HBA1*
	Epsilon	ε	*HBE1*
	^{G}Gamma	^{G}γ	*HBG2*
	^{A}Gamma	^{A}γ	*HBG1*
	Beta	β	*HBB*
	Delta	δ	*HBD*
Locus control genes	Locus control region, alpha (HS –40)	LCRα	*LCRA*
	Locus control region, beta	LCRβ	*LCRB*

subdivided into two peaks that were labelled, in order of their elution, haemoglobin A_I and haemoglobin A_{II} [4]; a little later it was found possible to subdivide the haemoglobin A_I peak into five smaller peaks, which were called haemoglobins A_I a, b, c, d and e in order of their elution [5]. It was later considered that haemoglobin A_{Ie} was a storage artefact. Haemoglobin A_I a, b and c are all glycated and may increase in diabetes mellitus whereas haemoglobin A_{Id} is an ageing peak due to glutathione combining with the cysteine residue at β93 [6], increasing with age of the haemolysate. The haemoglobin previously designated A_3 on electrophoresis was found to be of similar nature to the A_{Ia} and A_{Ib} peaks seen on cation exchange column chromatography [7] and also on high performance liquid chromatography (HPLC).

It was realized that confusion could be caused by using the designations haemoglobin A_2 and haemoglobin A_{II} for different types of haemoglobin and therefore haemoglobin A_{II} of column chromatography was renamed haemoglobin A_0. One consequence of the different separations and nomenclatures is that haemoglobin A on electrophoresis is equivalent to the sum of haemoglobin A_I and A_0 as measured by cation exchange chromatography and by most automated HPLC systems. All variant haemoglobins studied have been shown to have similar adducts to those of haemoglobin A; for instance, haemoglobin S has haemoglobin S_I and haemoglobin S_0. Haemoglobin F also separates into two peaks, but for a different reason. The main peak is called haemoglobin F_0 (it used to be called F_{II}) and the earlier, minor peak on HPLC is called F_I. Haemoglobin F_I is acetylated and is usually only present in sufficient quantities to be detected in neonatal samples.

Isoelectric focusing will also separate haemoglobin A into haemoglobin A_0 and haemoglobin A_I and haemoglobin F into haemoglobin F_0 and F_I. If the haemoglobin is from an old specimen and has become oxidized and methaemoglobin is present, then the methaemoglobin will also separate on isoelectric focusing and appear as several dark brown bands migrating cathodal to the parent haemoglobin. Bands due to ageing of the sample (probably glutathione adducts) are anodal to the parent haemoglobin (see technical notes to the atlas pages).

The normal haemoglobins having been named, as variant haemoglobins were discovered they were initially assigned letters of the alphabet. Sickle haemoglobin was initially called haemoglobin B, later changed to haemoglobin S. Subsequently letters were assigned in alphabetical order. Haemoglobin B_2, a variant of haemoglobin A_2 now designated A'_2, then haemoglobins C, D, E, G and so on. By the time the letter Q was reached (haemoglobin Q-India) it was clear that the number of letters in the alphabet would prove inadequate for the large numbers of variant haemoglobins being discovered and the convention was adopted that a haemoglobin would be named for the place of its discovery.

Mutations – what can go wrong?

Evolution led to the duplication and subsequent alteration of primordial genes, giving us the α and

Table 1.3 Types of mutations that can affect globin genes.

Type of mutation	Example
Gene duplication	Triple α
Gene deletion	Deletion of one or both α genes Deletion of LCR,α or LCR,β
Gene fusion	δβ fusion with loss of normal δ and β (haemoglobin Lepore) βδ fusion with retention of normal δ and β (haemoglobin anti-Lepore) α2α1 fusion with effective loss of one α gene
Point mutation within exon	β^S leading to sickle cell haemoglobin
Point mutation within intron	New splice site leading to β thalassaemia
Point mutation in enhancer	β thalassaemia
Small deletions	
Without frameshift	Haemoglobin Gun Hill (lacks five amino acids)
With frameshift	Haemoglobin Wayne (elongated α chain)
Deletion plus inversion	Indian type of deletional $^{A}\gamma\delta\beta^0$ thalassaemia
Deletion plus insertion	$--^{Med}\alpha^0$ thalassaemia
Insertion	
Without frameshift	Haemoglobin Grady (three extra amino acids in α chain)
With frameshift	Haemoglobin Tak (elongated β chain)

β clusters that are now part of the human genome. However, some mutations that have occurred during the course of evolution are potentially harmful. There may be an advantage for heterozygous carriers of mutant genes since some variant haemoglobins offer partial protection from the effects of malaria and the mutant gene therefore persists and its prevalence tends to increase; however in the homozygote (or compound heterozygote) the effects can be damaging. This is so for the sickle cell mutation and for the many mutations leading to β thalassaemia.

Mutations of globin genes are very varied in nature (Table 1.3). They include deletion or duplication of genes, formation of fusion genes, point mutations, small deletions within genes and deletions accompanied by inversions, and insertions, with or without an accompanying deletion. Gene deletion, gene duplication and formation of fusion genes can all result from unequal crossover during meiosis (the process by which germ cells are formed). Point mutations are very varied in their effects (Table 1.4). The genetic code is described as redundant, meaning that more than one triplet codon encodes the same amino acid. The consequence of this is that an alteration in the DNA sequence sometimes does not result in any alteration in the amino acid encoded. Other consequences of a point mutation range from a harmless substitution to one that has a severe clinical phenotype in homozygotes or compound heterozygotes or, occasionally, in simple heterozygotes. (The term 'compound heterozygote' refers to someone who has two different mutant alleles of a gene whereas a simple heterozygote has one normal and one abnormal allele.)

The consequences of small insertions and deletions and other more complex rearrangements (Table 1.3) are diverse. Deletion or insertion of three nucleotides or a multiple of three has effects rather similar to a point mutation since there is no alteration of the reading frame. However, the deletion or insertion of other numbers of nucleotides leads to a shift in the reading frame, which leads to all downstream triplets encoding different amino acids and also gives the possibility of creating a new STOP codon, with a resultant globin chain that is shortened as well as abnormal, or reading through the original STOP codon to give a chain that is elongated as well as abnormal.

Table 1.4 Some of the possible consequences of a point mutation within a globin gene.

Site of mutation	Possible effect of mutation	Example of functional consequence
Promoter	Reduced transcription	β^+ thalassaemia
Initiation codon	Methionine not encoded, absent transcription	β^0 thalassaemia
Exon	Same amino acid encoded (same-sense mutation)	None
	Different amino acid encoded (mis-sense mutation)	None Tendency to polymerize Tendency to crystallize Instability Tendency to oxidize, forming methaemoglobin High oxygen affinity Low oxygen affinity
	Coding sequence converted to STOP codon (non-sense mutation)	Shortened protein that may be very unstable or synthesised at a reduced rate
	Gene conversion	Often none (e.g. when $^{G}\gamma$ is converted to $^{A}\gamma$)
Splice site	Absence of normal transcription	β^0 thalassaemia
Consensus site	Reduced transcription	β^+ thalassaemia
Intron	False splice site created	β^+ thalassaemia
STOP codon	STOP codon converted to another STOP codon	No effect
	STOP codon converted to a coding sequence	Elongated globin chain, often synthesized at a reduced rate, α thalassaemia

The term 'haemoglobinopathy' is now usually used to indicate any abnormality of globin chain synthesis. When used in this sense, it encompasses a reduced rate of synthesis of one or more of the globin chains, a condition designated 'thalassaemia'. Most haemoglobinopathies are inherited or, much less often, result from mutation in a germ cell. However, there are occasional acquired haemoglobinopathies that result from somatic mutation, e.g. acquired haemoglobin H disease as a feature of a myelodysplastic syndrome.

This book deals with variant haemoglobins and some thalassaemias. All the common disorders and many rare disorders are included. It is directed at those working in diagnostic laboratories or seeking to interpret the results of investigations of globin chain disorders. It does not seek to cover the clinical aspects of haemoglobinopathies, although the laboratory results are interpreted in the context of the clinical significance of the disorder.

The proportion of variant haemoglobins

It might be expected that if one α gene were mutated the variant haemoglobin would be 25% of the total and that, similarly, if one β gene were mutated the variant haemoglobin would be 50% of the total. However, although this is true in general, the situation is far more complex. The proportion of a variant haemoglobin is influenced by: (i) whether it results from a mutation of an $\alpha 1$, $\alpha 2$, β, γ, δ or other gene; (ii) whether the variant chain is synthesized at a reduced rate; (iii) the charge of the variant chain (since this influences its affinity for the normal globin chain with which it forms a dimer); (iv) whether the variant globin chain or the resultant variant haemoglobin is unstable; (v) whether cells containing the variant haemoglobin survive normally; (vi) whether there is coexisting α or β thalassaemia; (vii) whether there are extra copies of the α globin gene; and (viii) acquired abnormalities such as iron deficiency.

References

1. Kunkel HG and Wallenius G (1955) New haemoglobin in normal adult blood. *Science*, **122**, 288.
2. Lehmann H (1957) News and views: International Society of Hematology, the haemoglobinopathies. *Blood*, **12**, 90–92.
3. Kunkel HG and Bearn AG (1957) Minor hemoglobin components of normal human blood. *Fed Proc*, **16**, 760–762.
4. Allen DW, Schroeder WA and Balog J (1958) Observations on the chromatographic heterogeneity of normal adult and fetal hemoglobin: a study of the effect of crystallization and chromatography on the heterogeneity and isoleucine content. *J Amer Chem Soc*, **80**, 1628–1634.
5. Clegg MD and Schroeder WA (1959) The chromatographic study of normal adult human hemoglobin including a comparison of hemoglobin from normal and phenylketonuric individuals. *J Amer Chem Soc*, **81**, 6065–6069.
6. Huisman THJ and Horton BF (1965) Studies of the heterogeneity of hemoglobin VII: chromatographic and electrophoretic investigations of minor hemoglobin fractions present in normal and *in vitro* modified red blood cell hemolysates. *J Chromat*, **18**, 116–123.
7. Schnek AG and Schroeder WA (1960) The relation between the minor components of whole normal human adult hemoglobin as isolated by chromatography and starch block electrophoresis. *J Amer Chem Soc*, **83**, 1472–1478.

Further reading

Bain BJ (2006) *Haemoglobinopathy Diagnosis*, 2nd edn. Blackwell Publishing, Oxford.

Globin Gene Server. http://globin.cse.psu.edu/. Electronic database hosted by Pennsylvania State University, USA and McMaster University, Canada (accessed 1 January 2010).

Steinberg MH, Forget BG, Higgs DR and Weatherall DJ (2009) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*, 2nd edn. Cambridge University Press, Cambridge.

Weatherall DJ and Clegg JB (2001) *The Thalassaemia Syndromes*, 4th edn. Blackwell Science, Oxford.