Haemoglobinopathy Diagnosis

Barbara J. Bain MBBS FRACP FRCPath

Professor of Diagnostic Haematology
St Mary's Hospital Campus of Imperial College
Faculty of Medicine, London
and Honorary Consultant Haematologist
St Mary's Hospital, London

Second Edition

Blackwell Publishing

Haemoglobinopathy Diagnosis

This book is dedicated to the past and present scientific staff of the Haematology Departments of the Princess Alexandra Hospital, Brisbane, Australia and St Mary's Hospital, Paddington, London, without whom it would not have been possible.

Haemoglobinopathy Diagnosis

Barbara J. Bain MBBS FRACP FRCPath

Professor of Diagnostic Haematology
St Mary's Hospital Campus of Imperial College
Faculty of Medicine, London
and Honorary Consultant Haematologist
St Mary's Hospital, London

Second Edition

Blackwell Publishing

Published by Blackwell Publishing Ltd
Blackwell Publishing, Inc., 350 Main Street, Malden, Massachusetts 02148-5020, USA
Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK
Blackwell Publishing Asia Pty Ltd, 550 Swanston Street, Carlton, Victoria 3053, Australia

First published 2001
Reprinted 2003
Second edition 2006

Library of Congress Cataloging-in-Publication Data

Bain, Barbara J.
Haemoglobinopathy diagnosis / Barbara J. Bain.—2nd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN-10: 1-4051-3516-6
ISBN-13: 978-1-4051-3516-0
1. Hemoglobinopathy—Diagnosis.
[DNLM: 1. Hemoglobinopathies—diagnosis. 2. Hematologic Tests—methods. WH 190 B162h 2006]
I. Title.

RC641.7.H35B73 2006
616.1'51075

2005015281

ISBN-13: 9-7814-0513-516-0
ISBN-10: 1-4051-3516-6

A catalogue record for this title is available from the British Library

Set in 9/12 Palatino by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in India by Replika Press Pvt., Ltd

Commissioning Editor: Maria Khan
Development Editor: Helen Harvey
Production Controller: Kate Charman

For further information on Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com

Contents

Preface to the first edition

It is now 6 years since colleagues at the Princess Alexandra Hospital, Brisbane, suggested that there was a need for a practical book on the laboratory diagnosis of haemoglobinopathies and that perhaps I might consider writing it. As the subject was one of considerable interest to me, I was happy to accept their suggestion. The book has been some time in the writing but here it is. I hope that it meets their expectations.

I am grateful to those with whom I have worked in this field of haematology. For over 25 years they have shared my pleasure in solving diagnostic problems and, at the same time, have joined in my efforts to provide an accurate, clinically relevant diagnostic service. It is to them that the book is dedicated.

I should like to thank particularly Ms Lorraine Phelan and Dr David Rees who have read the entire manuscript and have made many helpful suggestions. I am also grateful to the many colleagues who have contributed invaluable illustrations. They are individually acknowledged in the figure legends.

Barbara J. Bain
London, 2000

Abbreviations and glossary

α the Greek letter, alpha

α chain the α globin chain which is required for the synthesis of haemoglobins A, F and A_2 and also the embryonic haemoglobin, Gower 2

α gene one of a pair of genes on chromosome 16 that encode α globin

α thalassaemia a group of thalassaemias characterized by absent or reduced α globin chain transcription, usually resulting from the deletion of one or more of the α globin genes; less often it results from the altered structure of an α gene or the mutation of locus control genes or genes encoding *trans*-acting factors

α^0 thalassaemia a thalassaemic condition in which there is no α globin chain translation from one or both copies of chromosome 16

α^+ thalassaemia a thalassaemic condition in which there is reduced translation of α chain from one or both copies of chromosome 16

β the Greek letter, beta

β chain the β globin chain which forms part of haemoglobin A and haemoglobin Portland 2 and is the only globin chain in the abnormal haemoglobin, haemoglobin H

β gene the gene on chromosome 11 that encodes β globin

β thalassaemia a thalassaemia characterized by reduced β globin synthesis, usually caused by the mutation of a β globin gene; less often it results from gene deletion or from deletion or mutation of the locus control region

γ the Greek letter, gamma

γ chain the γ globin chain which forms part of fetal haemoglobin (haemoglobin F) and the embryonic haemoglobin, haemoglobin Portland 1, and is the only globin chain in the abnormal variant, haemoglobin Bart's

γ gene one of a pair of very similar genes on chromosome 11 encoding γ globin chain

γ thalassaemia a thalassaemic condition resulting from reduced synthesis of γ globin chain

δ the Greek letter, delta

δ chain a β-like globin chain, which forms part of haemoglobin A_2

δ gene a gene of the β cluster on chromosome 11 that encodes δ globin

δ thalassaemia a thalassaemic condition resulting from reduced synthesis of δ globin chain and therefore of haemoglobin A_2

ε the Greek letter, epsilon

ε chain the ε globin chain which is synthesized during early embryonic life and forms part of haemoglobins Gower 1 and Gower 2

ε gene a gene of the α globin cluster on chromosome 16 that encodes ε globin chain

ζ the Greek letter, zeta

ζ chain the ζ globin chain which is synthesized in intrauterine life and forms part of haemoglobins Gower 1, Portland 1 and Portland 2

ζ gene a gene of the α globin gene cluster on chromosome 16 that encodes ζ globin chain

ψ the Greek letter, psi, used to indicate a pseudogene

2,3-DPG 2,3-diphosphoglycerate; a small molecule that interacts with haemoglobin, decreasing its oxygen affinity

3′ the end of a gene where transcription ceases

5′ the end of a gene where transcription starts

acquired a condition that is not present at birth or is not inherited

affinity the avidity of haemoglobin for oxygen

ala δ-aminolaevulinic acid; the first compound formed during the process of haem synthesis

AML acute myeloid leukaemia

ARMS amplification refractory mutation system; a PCR technique used, for example, for the detection of mutations causing β thalassaemia; it employs two primer sets, one amplifying normal sequences and one abnormal sequences

balanced polymorphism the stable persistence of two or more alleles of a gene in a significant proportion of a population; a potentially deleterious allele may show balanced polymorphism if the heterozygous state conveys an advantage

base a ring-shaped organic molecule containing nitrogen which is a constituent of DNA and RNA; DNA contains four bases — adenine, guanine, cytosine and thymine; RNA contains four bases — adenine, guanine, cytosine and uracil

Bohr effect the effect of pH on oxygen affinity; the alkaline Bohr effect is the reduction of oxygen affinity of haemoglobin as the pH falls from above to below the physiological pH; there is also an acid Bohr effect which is a rise of oxygen affinity as the pH falls further, to a pH level that is incompatible with life

bp base pair; the pairing of specific bases, e.g. adenine with thymine, in the complementary strands of the DNA double helix

CAP 7-methylguanosine cap

carbonic anhydrase a red cell enzyme that is the second most abundant red cell protein after haemoglobin; it may be apparent on haemoglobin electrophoretic strips if a protein rather than a haem stain is used

carboxyhaemoglobin haemoglobin that has been chemically altered by combination with carbon monoxide

CE-HPLC cation-exchange high performance liquid chromatography

chromatography a method of separating proteins from each other by means of physical characteristics, such as molecular weight, charge or hydrophobicity, or by means of differing affinity for lectins, antibodies or other proteins; in column chromatography, the proteins move through an absorbent column and emerge after different periods of time

cis on the same chromosome (see also ***trans***)

***cis*-acting** a DNA sequence that affects the expression of a gene on the same chromosome but not on the homologous chromosome (see also ***trans*-acting**)

CO carbon monoxide, the molecule composed of one carbon atom and one oxygen atom, formed by combustion of hydrocarbons

CO_2 carbon dioxide, the molecule composed of one atom of carbon combined with two of oxygen

codon a triplet of nucleotides that encodes a specific amino acid or serves as a termination signal; there are 61 codons encoding 20 amino acids and three codons that act as termination or stop codons

congenital present at birth, often but not necessarily inherited

cooperativity the interaction between the four globin monomers that makes possible the Bohr effect and the sigmoid shape of the oxygen dissociation curve

CT computed tomography

CV coefficient of variation

DCIP 2,6-dichlorophenolindophenol

deletion loss of part of a chromosome, which may include all or part of a globin gene

deoxyhaemoglobin haemoglobin that is not combined with O_2

DGGE denaturing gradient gel electrophoresis; a molecular genetic technique for locating a mutation prior to precise analysis

DNA deoxyribonucleic acid; the major constituent of the nucleus of a cell; a polynucleotide strand that is able to replicate and that codes for the majority of proteins synthesized by the cell; the DNA molecule is a double helix of two complementary intertwined polynucleotides

EDTA ethylene diamine tetra-acetic acid

EKLF erythroid Kruppel-like factor

electrophoresis separation of charged suspended particles, such as proteins, by application to a membrane followed by exposure to a charge gradient, e.g. haemoglobin electrophoresis

ELISA enzyme-linked immunosorbent assay

elution removal of an absorbed substance from a chromatography column

enhancer a DNA sequence that influences the promoter of a nearby gene to increase transcription; an enhancer acts on a gene in *cis* and may be sited upstream, downstream or within a gene

exon a part of a gene that is represented in mature mRNA; most genes are composed of exons and non-translated introns

FAB classification French–American–British classification

FBC full blood count

Fe iron

Fe^{2+} ferrous or bivalent iron

Fe^{3+} ferric or trivalent iron

fetal haemoglobin haemoglobin F, the major haemoglobin present during intrauterine life, having two α chains and two γ chains

GAP-PCR a PCR technique in which there is amplification across a 'gap' created by deletion

GATA 1 an erythroid-specific transcription factor

gene the segment of DNA that is involved in producing a polypeptide chain; it includes regions preceding and following the coding region (5′ and 3′ untranslated regions) as well as intervening sequences (introns) between individual coding segments (exons); genes mediate inheritance; they are located on nuclear chromosomes or, rarely, in a mitochondrion

genetic code the relationship between a triplet of bases, called a codon, and the amino acid that it encodes

genotype the genetic constitution of an individual (cf. phenotype)

globin the protein part of the haemoglobin molecule, usually composed of two pairs of non-identical chains, e.g. two α chains and two β chains

H^+ a proton

haem a porphyrin structure that contains iron and that forms part of the haemoglobin molecule

haemoglobin a complex molecule composed of four globin chains, each one enclosing a haem group

haemoglobin A the major haemoglobin component present in most adults, having two α and two β chains

haemoglobin A_{1c} glycosylated haemoglobin A

haemoglobin A_2 a minor haemoglobin component present in most adults, and as an even lower proportion of total haemoglobin in neonates and infants, having two α chains and two δ chains

haemoglobin Bart's an abnormal haemoglobin with four γ chains and no α chains, present as the major haemoglobin component in haemoglobin Bart's hydrops fetalis and as a minor component in neonates with haemoglobin H disease or α thalassaemia trait

haemoglobin Bart's hydrops fetalis a fatal condition of a fetus or neonate with no α genes and, consequently, no production of haemoglobins A, A_2 or F

haemoglobin C a variant haemoglobin with an amino acid substitution in the β chain, mainly found in those of African ancestry

haemoglobin Constant Spring a variant haemoglobin with a structurally abnormal α chain which is synthesized at a reduced rate, leading to α thalassaemia

haemoglobin D the designation of a group of haemoglobin variants, some α chain variants and some β chain variants, that have the same mobility as haemoglobin S on electrophoresis at alkaline pH

haemoglobin dissociation curve a plot of the percentage saturation of haemoglobin against the partial pressure of oxygen

haemoglobin E a variant haemoglobin with an amino acid substitution in the β chain, mainly found in South-East Asia and parts of the Indian subcontinent

haemoglobin F fetal haemoglobin, the major haemoglobin of the fetus and neonate, which is present as a very minor component in most adults and as a larger proportion in a minority

haemoglobin G the designation of a group of haemoglobin variants, some α chain variants and some β chain variants, that have the same mobility as haemoglobin S on electrophoresis at alkaline pH

haemoglobin Gower 1 an embryonic haemoglobin, having two ε chains and two ζ chains

haemoglobin Gower 2 an embryonic haemoglobin, having two α chains and two ε chains

haemoglobin H a variant haemoglobin with four β chains and no α chains, present in haemoglobin H disease and, in small quantities, in α thalassaemia trait

haemoglobin H disease a haemoglobinopathy caused by marked underproduction of α chains, often consequent on deletion of three of the four α genes

haemoglobin I a group of variant haemoglobins that move more rapidly than haemoglobin A on electrophoresis at alkaline pH

haemoglobin J a group of variant haemoglobins that move more rapidly than haemoglobin A, but more slowly than haemoglobin I, on electrophoresis at alkaline pH

haemoglobin K a group of variant haemoglobins moving between haemoglobins A and J on electrophoresis at alkaline pH

haemoglobin Lepore a number of variant haemo-

globins resulting from the fusion of part of a δ globin gene with part of a β globin gene, giving a δβ fusion gene and fusion protein that combines with α globin to form haemoglobin Lepore

haemoglobin M a variant haemoglobin that oxidizes readily to methaemoglobin

haemoglobin N a group of variant haemoglobins moving between haemoglobins J and I on electrophoresis at alkaline pH

haemoglobin O-Arab a β chain variant haemoglobin moving near haemoglobin C at alkaline pH and near haemoglobin S at acid pH

haemoglobinopathy an inherited disorder resulting from the synthesis of a structurally abnormal haemoglobin; the term can also be used to encompass the thalassaemias in which there is a reduced rate of synthesis of one of the globin chains

haemoglobin Portland 1 an embryonic haemoglobin, having two ζ chains and two γ chains

haemoglobin Portland 2 an embryonic haemoglobin, having two ζ chains and two β chains

haemoglobin S sickle cell haemoglobin, a variant haemoglobin with a tendency to polymerize at low oxygen tension, causing erythrocytes to deform into the shape of a sickle

Hct haematocrit

HDW haemoglobin distribution width

heteroduplex analysis a molecular genetic technique for locating a mutation prior to precise analysis

heterozygosity the state of having two different alleles of a specified autosomal gene (or, in a female, two different alleles of an X chromosomal gene)

heterozygous having two different alleles of a specified autosomal or X chromosome gene

HIV human immunodeficiency virus

homologous being equivalent or similar to another

homologue an equivalent or similar structure; the α1 and α2 genes are homologues, as are the two copies of a chromosome

homology the presence of structural similarity, implying a common remote origin; the δ and β genes show partial homology

homozygosity the state of having two identical alleles of a specified autosomal gene

homozygous having two identical alleles of a specified autosomal gene (or, in a female, two identical alleles of an X chromosome gene)

HPLC high performance liquid chromatography; a method of separating proteins, such as haemoglobin variants, from each other on the basis of characteristics such as size, hydrophobicity and ionic strength; a solution of proteins is eluted from a specially designed column by exposure to various buffers, different proteins emerging after varying periods of time

HS1, HS2, HS3, HS4 hypersensitive sites 1, 2, 3 and 4

HS –40 an upstream enhancer of α globin gene transcription

HVR hypervariable region

IEF isoelectric focusing; the separation of proteins in an electric field as they move through a pH gradient to their isoelectric points

inherited a characteristic that is transmitted from a parent by means of genes that form part of nuclear or mitochondrial DNA

initiation (i) the process by which RNA transcription from a gene commences; (ii) the process by which protein translation from mRNA commences

initiation codon the three nucleotide codon (ATG) at the 5′ end of a gene which is essential to permit initiation of transcription of a gene, i.e. initiation of polypeptide synthesis

insertion the insertion of a DNA sequence, e.g. from one chromosome into another

intron a sequence of DNA in a gene which is not represented in processed mRNA or in the protein product

inversion the reversal of the normal position of a DNA sequence on a chromosome

isoelectric point the pH at which a protein has no net charge

IVS intervening sequence; an intron

kb kilobase; a unit for measuring the length of DNA; one kilobase is 1000 nucleotide base pairs

kDa kilodalton; a unit for measuring molecular weight; one kilodalton is 1000 daltons

LCR locus control region; a DNA sequence upstream of genes of the β globin cluster that enhances transcription of the genes of this cluster

LDH lactate dehydrogenase

MCH mean cell haemoglobin

MCHC mean cell haemoglobin concentration

MCV mean cell volume

methaemoglobin oxidized haemoglobin which does not function in oxygen transport

mis-sense mutation a mutation that leads to the encoding of a different amino acid

mRNA messenger ribonucleic acid; ribonucleic acid that is transcribed in the nucleus, on a DNA template, and moves to the cytoplasm, becoming attached to ribosomes and serving as a template for the synthesis of proteins

NO nitric oxide

nonsense mutation a mutation that leads to no amino acid being encoded and therefore functions as a stop or termination codon, leading to the synthesis of a truncated polypeptide chain

NRBC nucleated red blood cell

O_2 oxygen

ORF open reading frame

oxyhaemoglobin haemoglobin combined with O_2

P_{50} Po_2 at which haemoglobin is 50% saturated

Pao_2 partial pressure of oxygen in arterial blood

partial pressure of oxygen that part of the total blood gas pressure exerted by oxygen

PAS periodic acid–Schiff

PCR polymerase chain reaction; a method of making multiple copies of a DNA sequence

PCV packed cell volume

phenocopy a condition that simulates an inherited condition; a phenocopy may be acquired or may be a genetic characteristic that simulates another

phenotype the characteristics of an individual, which may be determined by the genotype, or may be an acquired characteristic (cf. genotype)

Po_2 partial pressure of oxygen

polymorphism the occurrence of a variant form of a gene in a significant proportion of a population

promoter a sequence of DNA at the 5′ end of a gene which is essential for initiation of transcription

pseudogene a non-functioning homologue of a gene

purine one of the two types of nitrogenous base found in nucleic acids; purines have a double ring structure (see also **pyrimidine**)

pyrimidine one of the two types of nitrogenous base found in nucleic acids; pyrimidines have a single ring structure (see also **purine**)

RBC red blood cell count

RDW red cell distribution width

restriction endonuclease an enzyme that recognizes specific sequences in a DNA molecule and cleaves the molecule in or very near the recognition site

restriction fragment a fragment of DNA produced by cleavage by a restriction endonuclease

RFLP restriction fragment length polymorphism; variation between homologous chromosomes with regard to the length of DNA fragments produced by application of a specific restriction endonuclease; can be used for the demonstration of heterozygosity or for the demonstration of a specific gene that removes or creates a specific cleavage site

ribosome a cytoplasmic structure on which proteins are translated from mRNA; ribosomes may be free within the cytosol or form part of the rough endoplasmic reticulum

RNA ribonucleic acid; a polynucleotide in which the nitrogenous bases are adenine, guanine, cytosine and uracil and the sugar is ribose; RNA is produced in the nucleus and in mitochondria from DNA templates

rRNA ribosomal RNA; RNA that, together with protein, constitutes the ribosomes

sickle cell an erythrocyte that becomes sickle- or crescent-shaped as a result of polymerization of haemoglobin S

sickle cell anaemia the disease resulting from homozygosity for haemoglobin S

sickle cell disease a group of diseases including sickle cell anaemia and various compound heterozygous states in which clinicopathological effects occur as a result of sickle cell formation

sickle cell trait heterozygosity for the β^s gene that encodes the β chain of haemoglobin S

SOP standard operating procedure

splicing the process by which RNA sequences, corresponding to introns in the gene, are removed during processing of RNA

SSP stage selector protein

sulphaemoglobin haemoglobin that has been irreversibly oxidized and chemically altered by drugs or chemicals with incorporation of a sulphur atom into the haemoglobin molecule

thalassaemia a disorder, usually inherited, in which one or more of the globin chains incorporated into a haemoglobin molecule or molecules is synthesized at a reduced rate

thalassaemia intermedia a thalassaemic condition that is moderately severe, but nevertheless does not require regular blood transfusions to sustain life

thalassaemia major thalassaemia that is incompatible with more than a short survival in the absence of blood transfusion

thalassaemia minor an asymptomatic thalassaemic condition

trait a term applied to heterozygosity for an inherited characteristic; in the case of disorders of globin genes, the term would not be used if heterozygosity were associated with a significant phenotypic abnormality; rather it is used when homozygosity or compound heterozygosity produces a clinically significant abnormality but simple heterozygosity does not

trans having an influence on a DNA sequence on another chromosome (see also ***cis***)

***trans*-acting** a DNA sequence that affects the expression of a gene on another chromosome (see also ***cis*-acting**)

transcript an RNA molecule, corresponding to one gene, transcribed from nuclear DNA

transcription the synthesis of RNA on a DNA template

transcription factor a protein capable of enhancing transcription of one or more genes

translation the synthesis of protein from an mRNA template

tRNA transfer RNA; RNA molecules that bind to specific amino acids and transport them to ribosomes; there they bind to specific mRNA sequences, leading to incorporation of amino acids into peptide chains in the sequence specified by the mRNA

unstable a term applied to a haemoglobin that is abnormally prone to post-translational structural alteration, which may include loss of the normal tertiary or quaternary structure

UTR untranslated region

variant a term applied to any haemoglobin other than haemoglobins A, A_2, F and the normal embryonal haemoglobins

WBC white blood cell count

yolk sac part of an embryo; the initial site of formation of blood cells

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 is composed of two dissimilar pairs of polypeptide chains, each of which encloses an iron-containing porphyrin designated haem (Fig. 1.1). Haemoglobin has a molecular weight of 64–64.5 kDa. 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 modification. 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 reversible reaction, converting haem to haematin and haemoglobin to methaemoglobin, a form of haemoglobin that cannot transport oxygen.

The haemoglobin molecule can also combine with CO_2, being responsible for about 10% of the transport of CO_2 from the tissues to the lungs; transport is by reversible carbamation of the N-terminal groups of the α chains of haemoglobin. Carbamated haemoglobin has a lower oxygen affinity than the non-carbamated form, so that 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.

The haemoglobin molecule has a role in nitric oxide (NO) transport and metabolism. Haemoglobin is both a scavenger and an active transporter of NO. NO is produced in endothelial cells and neutrophils by the action of nitric acid synthase [1–3]. NO has a very high affinity for oxyhaemoglobin, so that blood levels are a balance between production and removal by binding to oxyhaemoglobin. NO is a potent vasodilator, this effect being 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 α chain) binds NO. A haemoglobin molecule with NO bound to two haem groups strikingly favours the deoxy conformation, so that oxygen is readily released. NO-haemoglobin is subsequently converted to methaemoglobin with release of NO and the production of nitrate ions, which are excreted. As deoxyhaemoglobin has a much lower affinity for NO, hypoxic conditions may leave more NO free and lead to vasodilation of potential physiological benefit.

NO 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 NO may be retained, with NO 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

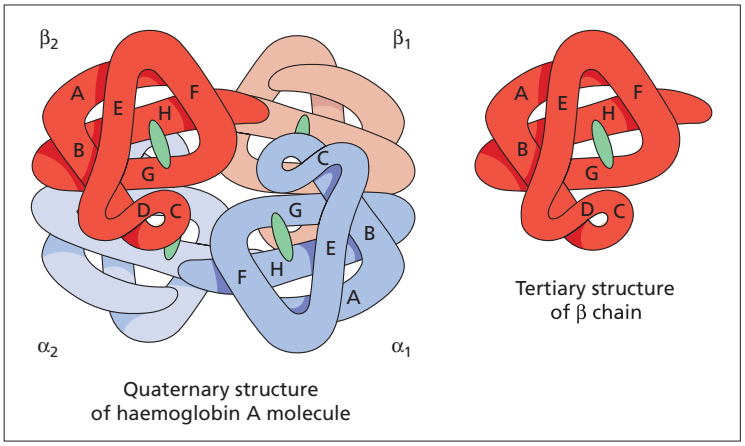

Fig. 1.1 Diagrammatic representation of the tertiary structure of a haemoglobin monomer (a β globin chain containing a haem group) and the quaternary structure of haemoglobin; upper case letters indicate homologous α helixes.

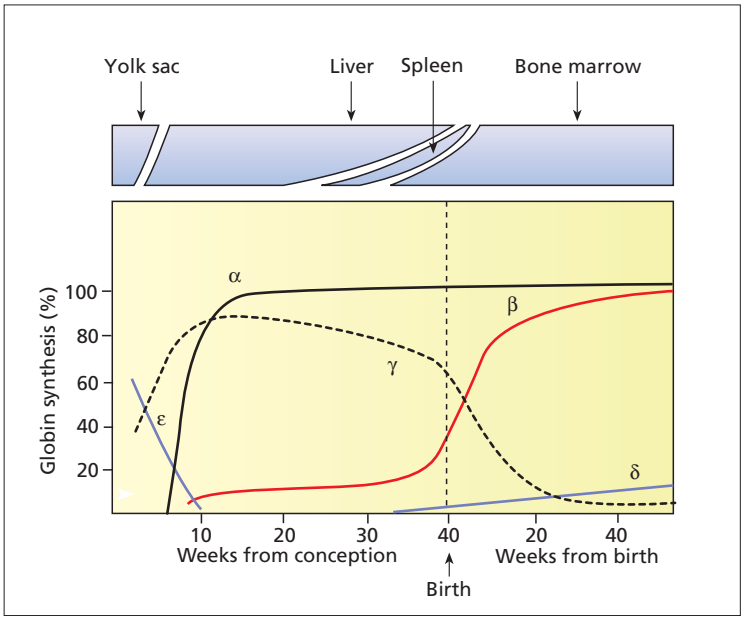

Fig. 1.2 Diagrammatic representation of the sites and rates of synthesis of different globin chains *in utero* and during infancy.

the release of NO. This may be an important physiological process, with NO 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 NO is reduced by the barrier to NO diffusion provided by the red cell membrane; however, in chronic haemolytic anaemia, increased free plasma haemoglobin may lead to impaired vascular responses to NO [3]; inactivation of NO by haemoglobin in the plasma may contribute to the pulmonary hypertension that can be a feature of sickle cell anaemia, and also the hypertension that has been observed with some haemoglobin-based blood substitutes.

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

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 fetal and neonatal life
F	$\alpha_2{}^G\gamma_2$ or $\alpha_2{}^A\gamma_2$	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.

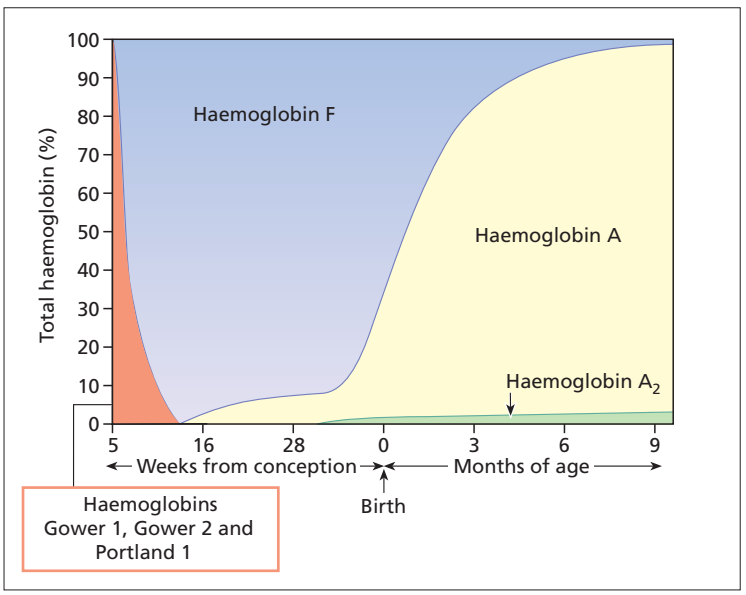

Fig. 1.3 Diagrammatic representation of the average percentages of various haemoglobins present during the embryonic and fetal periods and during infancy.

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 [4]. A minor haemoglobin, haemoglobin A_2, has two α chains and two delta (δ) chains. Its existence was first reported in 1955 by Kunkel and Wallenius [5]. 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 [6], a polymorphism that was previously referred to as haemoglobin F-Sardinia. In the early embryo, haemoglobin is synthesized 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 synthesized in significant amounts only during embryonic life, specifically zeta (ζ) and epsilon (ε) chains (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 [7], being named for Gower Street, 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 [8]. By 5 weeks of gestation, ζ and ε chains are already being synthesized in primitive erythroblasts in the yolk sac. From the sixth week onwards, these same cells start to synthesize α, β and γ chains. Starting from about the 10th–12th week of gestation, there is haemoglobin synthesis in the liver and the spleen with the 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 synthesize haemoglobin A and the minor haemoglobins.

The embryonic haemoglobins have a higher oxygen affinity than haemoglobin A, similar to that of haemoglobin F [9]. They differ from haemoglobins A and F in that they continue to bind oxygen strongly, even in acidotic conditions [9]. In the case of Gower 2, impaired binding to 2,3-diphosphoglycerate (2,3-DPG) is the basis of the increased oxygen affinity [10].

Haemoglobin can undergo post-translational modification (see also Chapter 6). Glycosylation occurs with the formation of haemoglobins A_{1a-e}, but principally haemoglobin A_{1c}. In normal individuals, haemoglobin A_{1c} may constitute up to 4–6% of total haemoglobin, but in diabetics it can be much higher. In individuals with a shortened red cell life span it 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 [6]. 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 sulphaemoglobin and methaemoglobin are also formed in normal subjects. Methaemoglobin is usually less than 1% of total haemoglobin. 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 below). In addition, some abnormal haemoglobins, in which there is a mutation of N-terminal amino acids, are particularly prone to acetylation, which occurs cotranslationally [11].

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 which constitutes the globin chain.

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

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

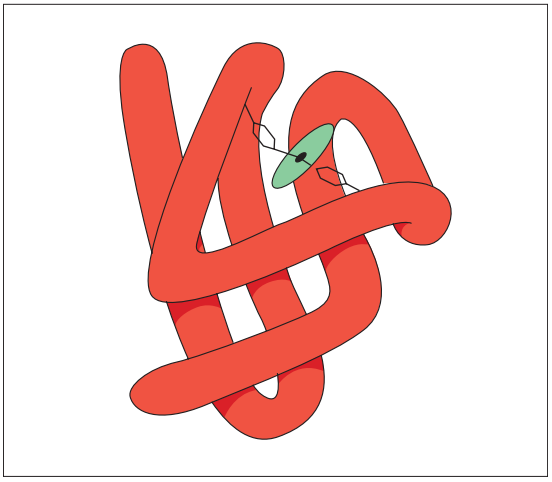

Fig. 1.4 Diagrammatic representation of a haemoglobin molecule with a haem group within the haem pocket, showing the relationship of haem to two histidine residues of the globin chain (designated proximal and distal histidines).

dissociation curve, the Bohr effect and the variation of oxygen affinity consequent on interaction with 2,3-DPG (see below). Contacts between like chains, $\alpha_1\alpha_2$ and $\beta_1\beta_2$, are also of physiological significance.

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

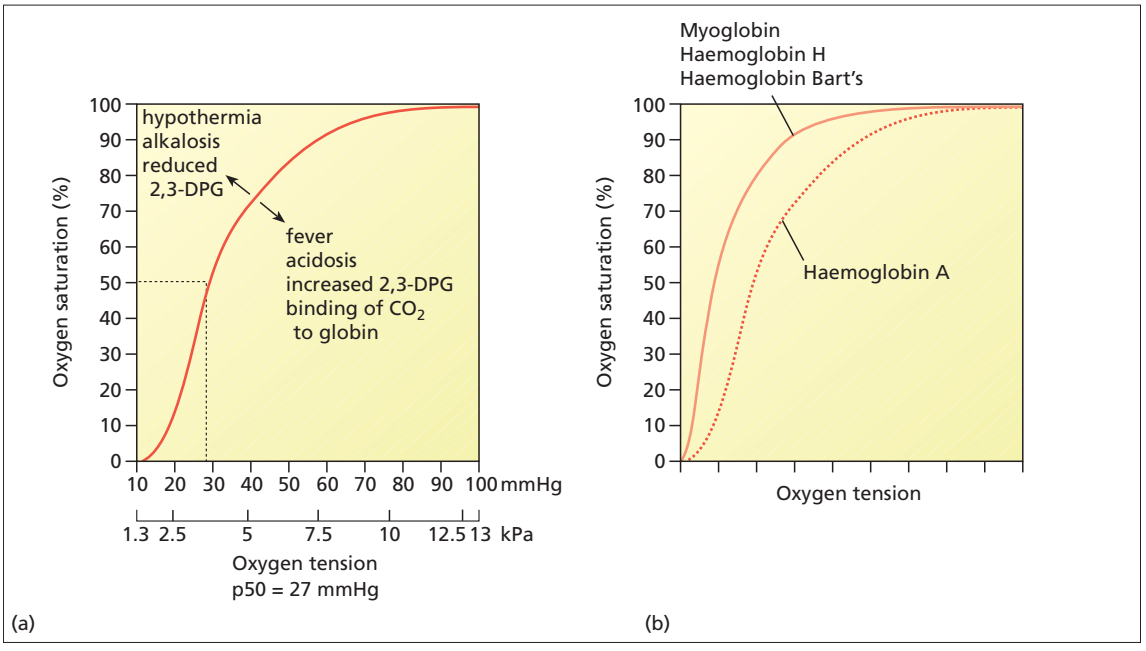

Fig. 1.5 (a) Normal oxygen dissociation curve indicating the effects of alteration of pH, body temperature and 2,3-diphosphoglycerate (2,3-DPG) concentration on the oxygen affinity of haemoglobin. (b) Comparison of the hyperbolic oxygen dissociation curve characteristic of myoglobin and of abnormal haemoglobins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

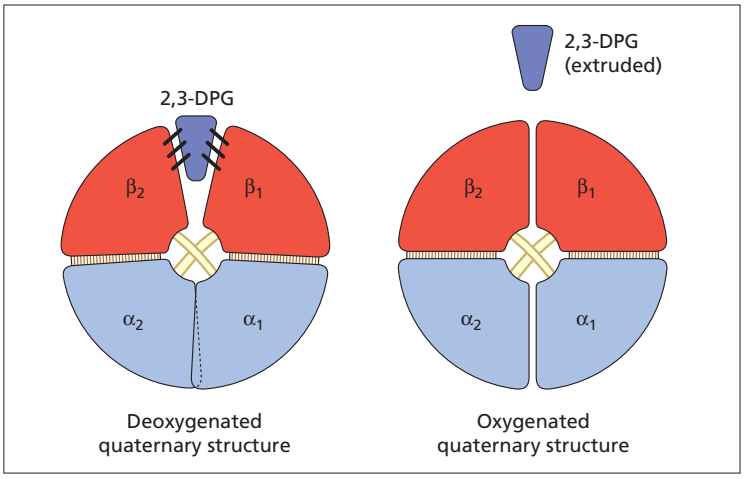

Fig. 1.6 Diagrammatic representation of the effect of oxygenation and deoxygenation on the quaternary structure of haemoglobin. The haemoglobin dimers ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) are stable, with the dimeric bonds between the α and β chains having 34 contacts in both the deoxygenated and oxygenated forms. There are less strong $\alpha_2\beta_1$ and $\alpha_1\beta_2$ tetrameric bonds, with 17 contacts between the α and β chains 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-diphosphoglycerate (2,3-DPG) binds to the β chains (three 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.

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 in oxygen affinity that occurs when the pH falls from physiological levels of 7.35–7.45 towards 6.0. Increasing metabolism in tissues lowers the pH as there is increased production of CO_2 and carbonic acid and, in addition, in anaerobic conditions, the 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 of these effects are likely to be of physiological significance. Fever increases the metabolic rate, so that decreased oxygen affinity, favouring downloading of oxygen, is beneficial in this circumstance. The lower pH in tissues favours the delivery of oxygen to sites of active metabolism, whereas the efflux of CO_2 in the lungs raises the pH and favours the uptake of oxygen by haemoglobin. It should 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, under separate genetic control, are known to be necessary for haem synthesis [12]. 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 occur

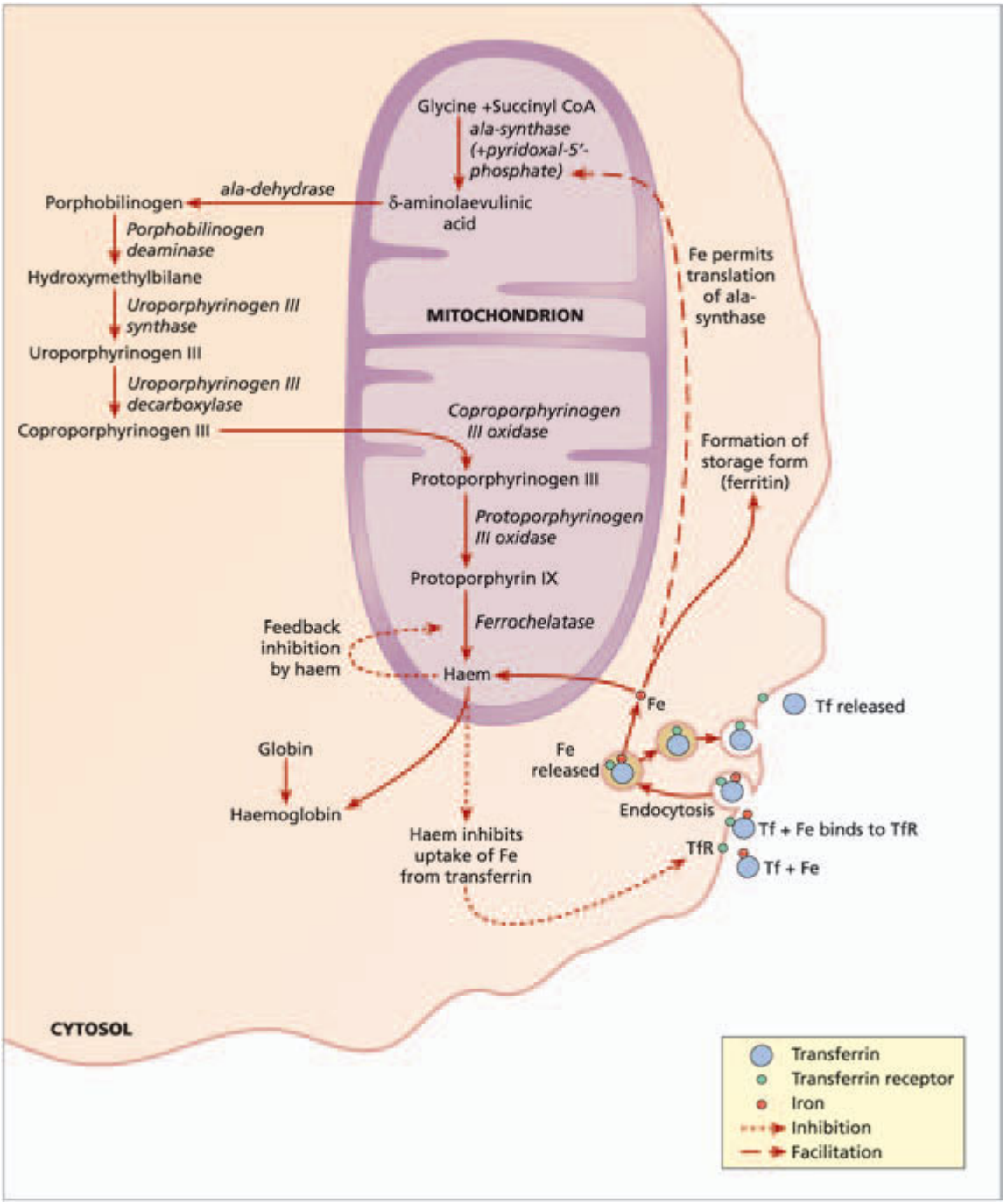

Fig. 1.7 Diagrammatic representation of haem synthesis.

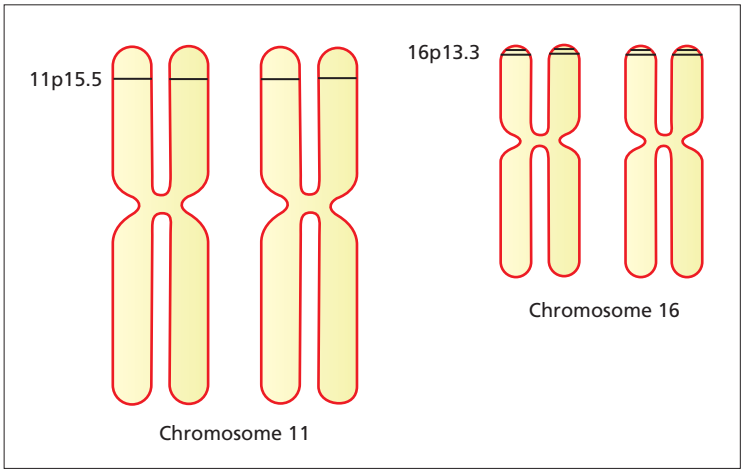

Fig. 1.8 Diagram of chromosomes 11 and 16 showing the positions of the β and α globin gene clusters.

in the mitochondrion, whereas the four intermediate enzymatic reactions occur in the cytosol. The first rate-limiting step in haem synthesis is the formation of δ-aminolaevulinic acid (ala) by condensation of glycine and succinyl CoA. This reaction is under the control of ala-synthase with pyridoxal-5′-phosphate as cofactor. The rate of formation of ala is controlled by iron availability; iron deficiency causes iron regulatory proteins to bind to iron-responsive elements in the messenger RNA (mRNA) for ala-synthase with consequent repression of translation. Synthesis of ala is followed by its entry into the cytosol where two molecules combine, under the influence of 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 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.

The uptake of iron by erythroid cells is from transferrin (Fig. 1.7). A molecule of transferrin with its attached iron first binds to a membrane transferrin receptor. The whole complex is internalized, a process known as endocytosis. Iron is released from its carrier within the endocytotic vesicle and, following reduction to the ferrous form, is 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 inhibits both ferrochelatase and the acquisition of iron from transferrin. Reduced cellular uptake of iron in turn inhibits the production of ala. Uptake of iron by erythroid cells is enhanced by iron deficiency and by increased levels of erythropoietin. Both lead to the 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.

The synthesis of α and β globin chains takes place in erythroid precursors, from the proerythroblast to the reticulocyte stage; δ chain synthesis ceases before the reticulocyte stage [13]. Synthesis is in the cytoplasm, on ribosomes. The genes for 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), a kilobase being 1000 nucleotide bases. The β gene is at 11p15.5. 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,

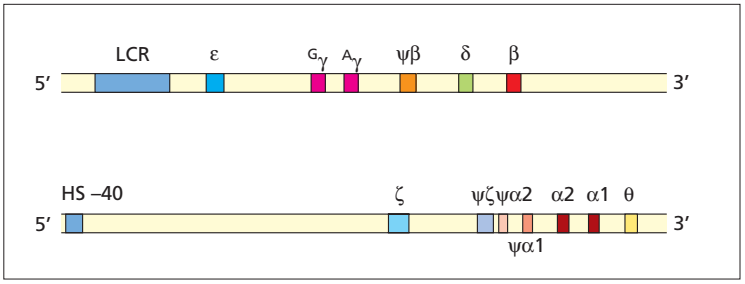

Fig. 1.9 Diagrammatic representation of the α and β globin gene clusters.

Table 1.2 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		**CCAAT**	**TATA**AAC
α1 and α2		**CCAAT**	C**ATA**AAC
ε		**CCAAT**	A**ATA**AAG
^{G}γ and ^{A}γ	**CACCC**	**CCAAT/CCAAT**	A**ATA**AAA
β	**CACCC**	**CCAAT**	C**ATA**AAA
δ		**CCAA**C	C**ATA**AAA

a ζ gene (also referred to as ζ2), a pseudo ζ gene (ψζ or ψζ1), two pseudo α genes, (ψα2 and ψα1) and two α genes, designated α2 and α1. The β cluster on chromosome 11 contains, in the following order, an ε gene, two γ genes, designated ^{G}γ and ^{A}γ, respectively, a pseudo β gene (ψβ), a δ gene and a β gene. 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 is 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 α and β genes 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, a sequence that binds RNA polymerase and transcription factors and is necessary for the initiation of transcription. Globin gene promoters share several conserved DNA sequences that bind crucial transcription factors [14,15]. These are summarized in Table 1.2.

The process by which globin chains are synthesized is shown diagrammatically in Fig. 1.10. Transcription is the process by which RNA is synthesized 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 to both promoters and upstream regulatory elements, referred to as the β-locus control region (β-LCR) for the β cluster and HS –40 for the α cluster. The β-LCR includes four

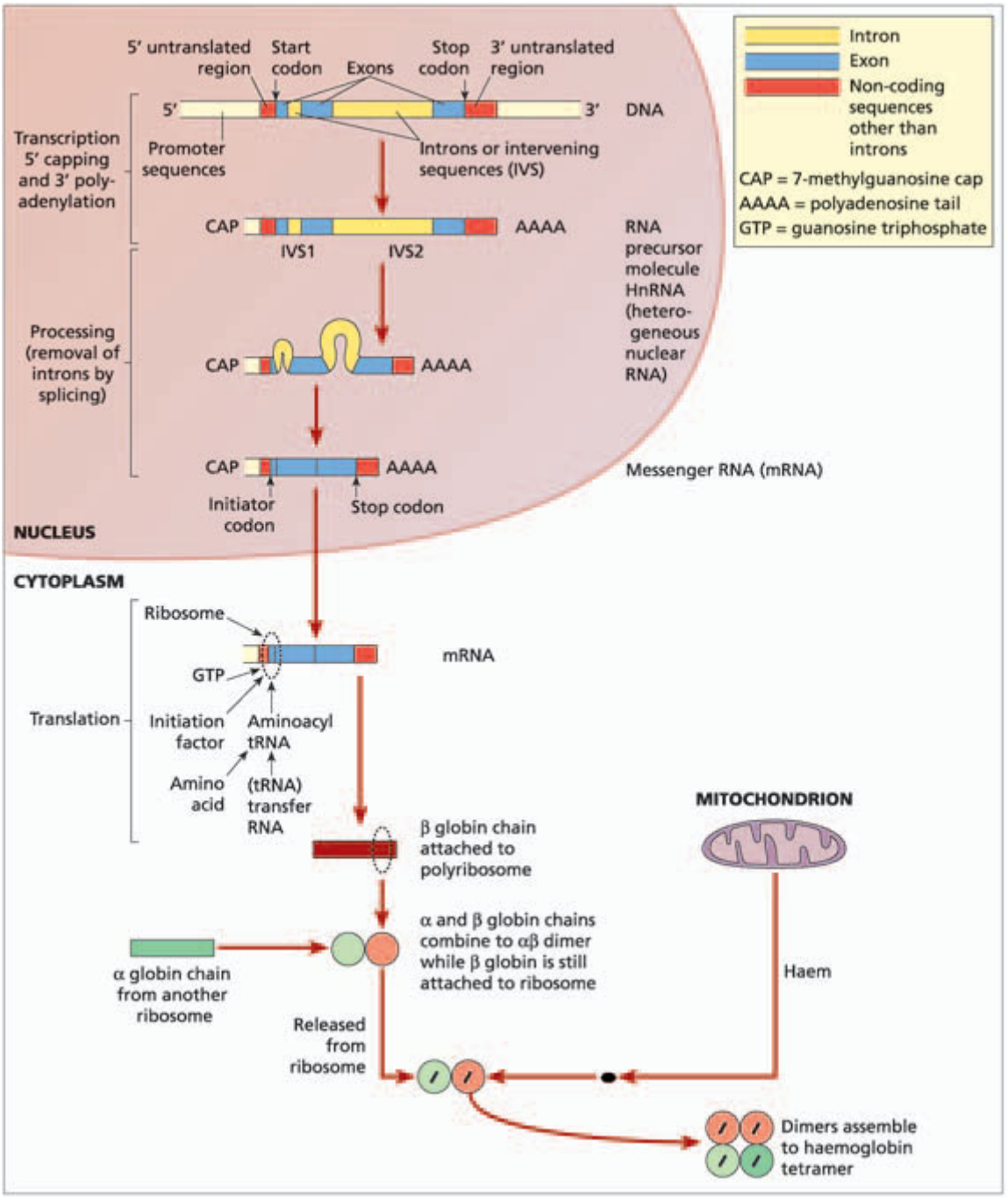

Fig. 1.10 Diagrammatic representation of RNA synthesis and processing and β globin chain synthesis.

erythroid-specific DNase sites, designated HS1, HS2, HS3 and HS4, 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 [16]. There are also enhancers 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 the expression of globin genes. Relatively erythroid-specific *trans*-activating factors, including GATA1, NFE2, EKLF, SSP, Nrf-1, Nrf-2 and LCR-F1, contribute to the regulation of gene expression by interacting either with the LCRs or with the globin gene promoters to increase gene expression [17,18]. EKLF (erythroid Kruppel-like factor) is an enhancer of β chain synthesis and SSP (stage selector protein) is an enhancer of δ and γ chain synthesis [17]. 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, Sp1, YY1, USF and TAL-1/SCL [16–18]. Nascent RNA molecules resulting from transcription are large and unstable and are modified in the nucleus. Initially, the 5′ end acquires a 7-methylguanosine cap (CAP), which is probably added during transcription and has a role during translation; during this 'capping' process, methylation of adjacent ribose residues also occurs. Following this, 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) that serves as a signal for 3′ cleaving of the transcript and polyadenylation. Polyadenylation may have a role in transfer of the mRNA from the nucleus to the cytoplasm. The polyadenylate tail is also important for mRNA stability and enhances translation. Finally, the introns are excised to give a functional mRNA molecule 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 for both a nucleotide triplet and an amino acid. Amino acids are thus assembled in the correct sequence, 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′ UTR and encodes methionine. Initiation requires the amino acid methionine, tRNA specific for methionine, guanosine triphosphate (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. When the chain reaches 40–50 residues, cotranslational acetylation of the N-terminal residue can occur through the action of several acetyl transferases [19]. 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. 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, i.e. the next step after initiation. Transcription from the two α genes is equal up to the eighth 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 [20]. The translational efficiency differs somewhat so that the α2 gene directs the synthesis of about twice as much α chain as the α1 gene. There is more α than β mRNA, probably about 2.5 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 [20]. The 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, the 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 synthesized on different polyribosomes. The combination of a free α

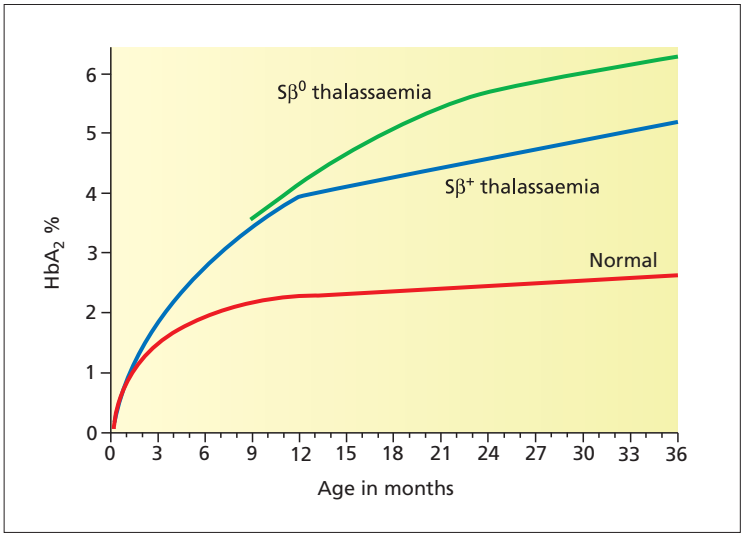

Fig. 1.11 Diagram showing the rate of rise of haemoglobin A_2 in haematologically normal Jamaican babies and in babies with sickle cell/β thalassaemia. (Modified from reference [21].)

chain with a β chain that is still attached to the polyribosome, to form an αβ dimer, may contribute to the 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. Both the α and β globin genes have structural determinants in their 3′ UTRs that are important for mRNA stability [17].

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 adults, haemoglobin A_2 comprises about 2–3.5% of total haemoglobin. The percentage is much lower at birth, about 0.2–0.3%, 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 [21] (Fig. 1.11). In the normal adult population, the percentage of haemoglobin A_2 shows a Gaussian distribution. It has functional properties that are very similar to those of haemoglobin A [13] (similar cooperativity and interaction with 2,3-DPG), although, in comparison with haemoglobin A, it inhibits polymerization of haemoglobin S [22] and has a higher oxygen affinity [10]. 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 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 [13] and, in addition, lacks the CACCC sequence that is present in the β promoter (Table 1.2). The proportion of haemoglobin A_2 is reduced by absolute or functional iron deficiency (see Table 6.3) and by α, δ and δβ thalassaemia trait (see Fig. 3.11). In γδβ thalassaemia, the rate of synthesis, but not the proportion, of haemoglobin A_2 is reduced, as the 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 and in some patients with an unstable haemoglobin.

There are δ chain variants and δ thalassaemias. About 1% of individuals of African ancestry have the variant haemoglobin designated haemoglobin A_2' (A_2 prime) or haemoglobin B_2 ($\delta^{16Gly\rightarrow Arg}$). It is readily detected by high performance liquid chromatography (Fig. 1.12) and isoelectric focusing. The δ thalas-

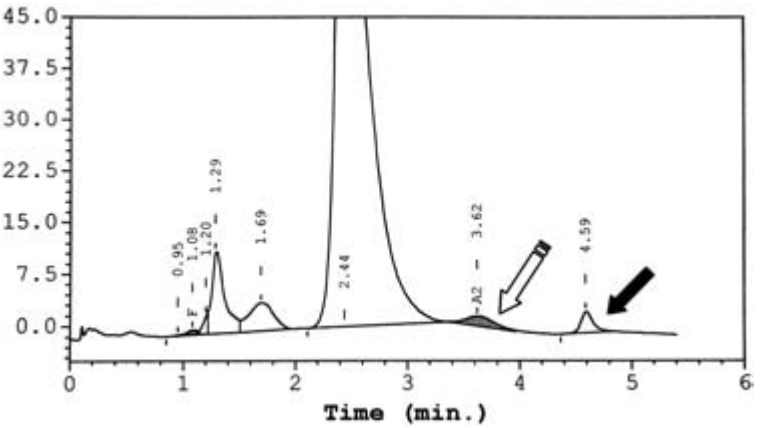

Fig. 1.12 High performance liquid chromatography (HPLC) chromatogram showing a split haemoglobin A_2 resulting from heterozygosity for haemoglobin A_2'; the white arrow shows haemoglobin A_2 and the black arrow haemoglobin A_2'.

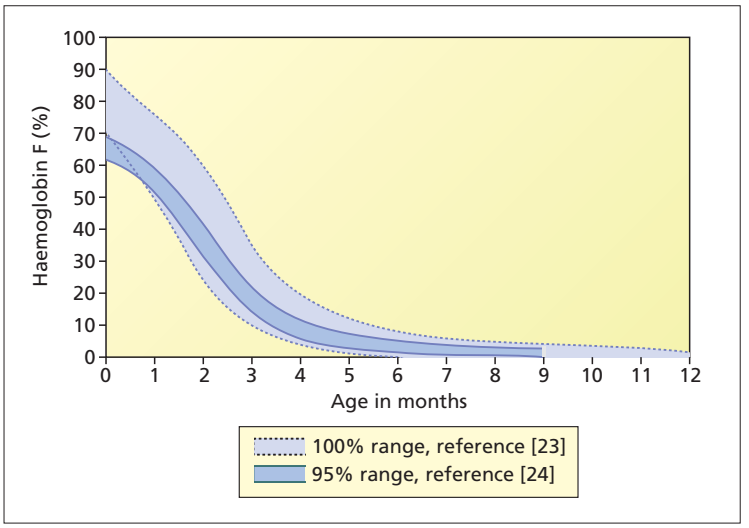

Fig. 1.13 Rate of fall of the percentage of haemoglobin F postnatally in normal and premature babies; the pale blue represents premature babies while the deep blue represents normal babies. (Derived from references [23,24].)

saemias are also common in some ethnic groups, e.g. present in 1% of Sardinians [10]. The δ thalassaemias and δ chain variants are of no functional significance, although some variants are unstable or have increased oxygen affinity. However, their presence complicates the diagnosis of β thalassaemia trait (see p. 97).

Haemoglobin F

Haemoglobin F is the major haemoglobin during intrauterine 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 appears to be normal in the offspring of mothers with very high levels of haemoglobin F. 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 [6]. 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 year of life, the percentage of haemoglobin F falls progressively to values close to adult levels (Fig. 1.13) [23–25]. 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 concentration lasting 20–60 days, followed by a linear decrease similar to that in term babies [24]. At any given period after birth, the spread of values is greater than that 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 [24].

In normal adults, haemoglobin F is heterogeneously distributed, being found in a subset of erythrocytes designated F cells. The proportion of F cells is highly variable, in one study ranging from 0.6% to 22% [26]. The percentage of haemoglobin F is determined by age, sex (slightly higher in women) and a number of inherited characteristics both linked and unlinked to the β globin gene cluster. DNA sequences controlling the proportion of F cells and the percentage of haemoglobin F include [18,23–29]:

- a polymorphism at position –158 of the Gγ gene (C→T being associated with a higher haemoglobin F);
- variation of the number of repeats of a specific motif at –530 in the HS2 component of the β-LCR, namely $(AT)_xN_{12}GT(AT)_y$;
- a *trans*-acting locus at 6q22.3–23.2;
- a *trans*-acting locus at Xp22.2–22.3;
- a *trans*-acting locus on an autosome other than 6.

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 LCR at –530 base pairs (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 BP-1, a negative *trans*-acting factor [30].

The distribution of the percentage of haemoglobin F in the population is skewed. In 85–90% of individuals, haemoglobin F is less than 0.6–0.7% and F cells are less than 4.5% [27,29]. 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 [27]. However, as 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 (see Table 3.12) and, less often, in various acquired conditions (see Table 6.2).

Variant haemoglobins and abnormalities of globin gene 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 may occur in globin genes are summarized in Table 1.3. 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).

Point mutations in globin genes sometimes have no effect on the amino acid sequence. This occurs because, as mentioned above, there is redundancy in the genetic code, with a number of nucleotide triplets coding for 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 mutations in globin genes are same-sense mutations. Point mutations may also result in a 'missense' mutation when the new codon codes for a different amino acid, leading to the production of a variant haemoglobin. The site of a mutation is critical, determining whether there is an effect on stability, oxygen affinity, solubility or other critical characteristics of the haemoglobin molecule. Because of the redundancy in the genetic code, different point mutations may give rise to the same variant

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

Type of mutation	Possible consequence	Example
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 one-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 CD116 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
	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 (ζ1) Conversion between the α2 and α1 genes so that the same mutation is present in both, e.g. $\alpha2^{Lys\rightarrow Glu}\alpha1^{Lys\rightarrow Glu}$, giving unusually high levels of haemoglobin I
	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
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

Continued on p. 16.

Table 1.3 *Continued.*

Type of mutation	Possible consequence	Example
Deletion or duplication of one or more genes		
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 or quadruple α gene	ααα†/αα, ααα/ααα, αααα‡/αα or αααα/αααα
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 eight 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 four ζ-like genes per chromosome
Abnormal cross-over during meiosis leading to gene fusion		
α2α1 fusion	Effective loss of one α gene but structurally normal α chain is encoded	$-\alpha^{3.7}$ thalassaemia
δβ fusion — simple cross-over	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 [31]
δβδ fusion — double cross-over 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
$^{A}\gamma\beta$ 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
$^{G}\gamma^{A}\gamma$ fusion (designated $-^{G}\gamma^{A}\gamma-$)	Reduced rate of synthesis of haemoglobin F	γ thalassaemia

Table 1.3 *Continued.*

Type of mutation	Possible consequence	Example
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 $^{A}\gamma$ 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)
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
Chromosomal translocation		
Unbalanced translocation	Extra α genes on a chromosome other than chromosome 16	Same significance as homozygous triplication of an α gene as there are a total of six α genes
	Loss of an α gene	α thalassaemia
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	$(\varepsilon)\gamma\delta\beta^{0}$ thalassaemia
	Deletion of the α gene enhancer (HS –40) 40 kb upstream of the ζ2 gen	α^{0} thalassaemia

* 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; it 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 $\alpha\alpha\alpha\alpha^{anti4.2}$.