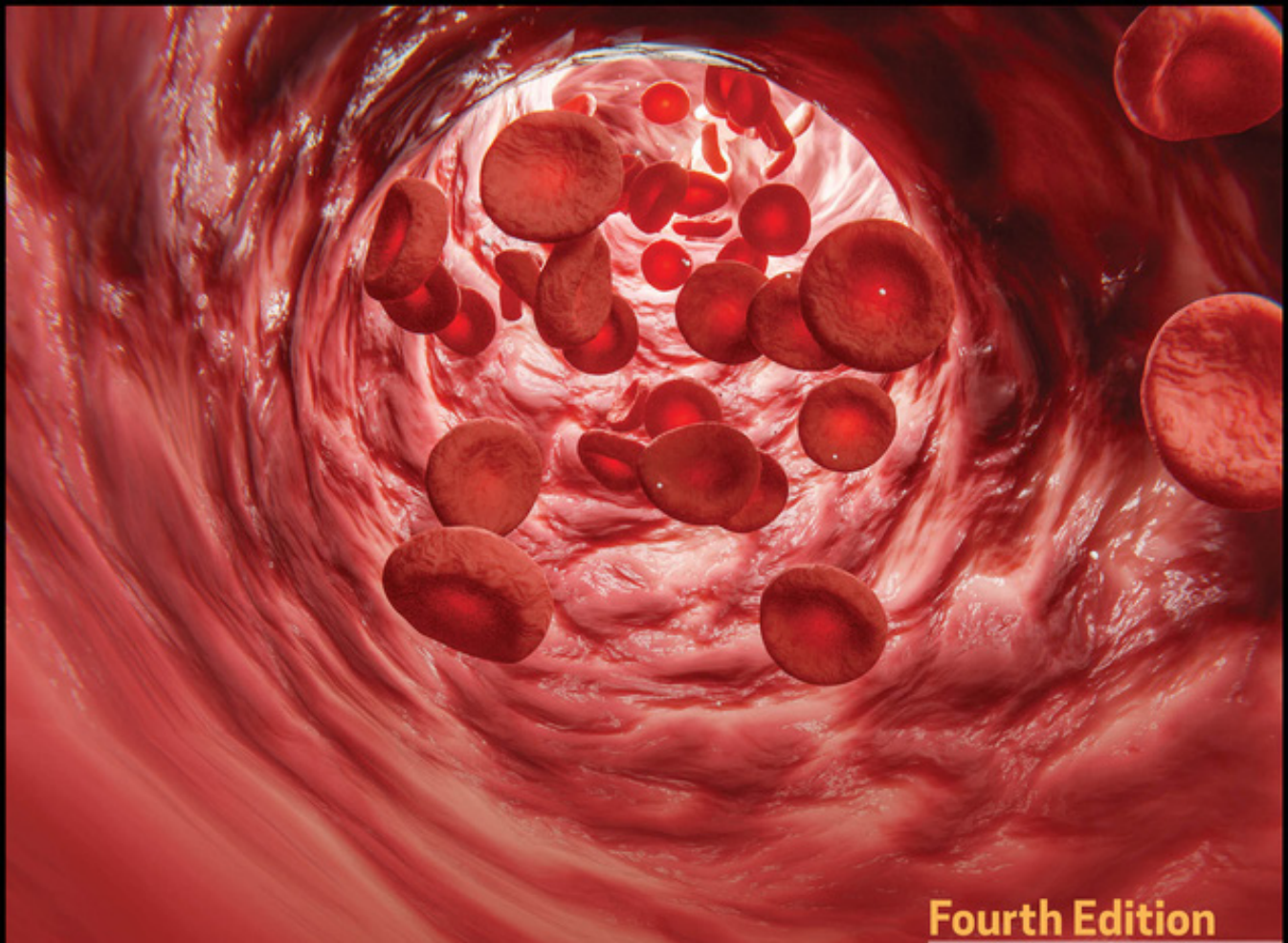
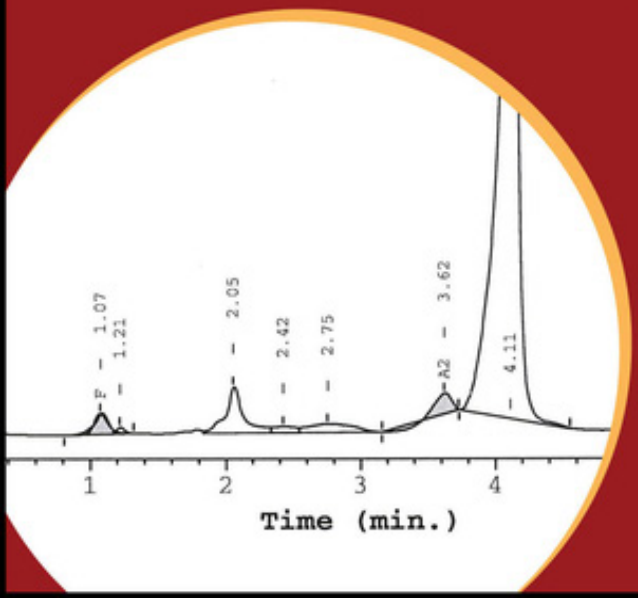


Fourth Edition

Haemoglobinopathy Diagnosis



Barbara J. Bain
David C. Rees

WILEY Blackwell

Haemoglobinopathy Diagnosis

Haemoglobinopathy Diagnosis

Barbara J. Bain MBBS FRACP FRCPath

Professor of Diagnostic Haematology,
St Mary's Hospital Campus of Imperial College, London, UK

Honorary Consultant Haematologist,
St Mary's Hospital, London, UK

David C. Rees MA MBBS FRCP FRCPath FRCPCH

Professor of Paediatric Haematology,
King's College London, UK

Honorary Consultant in Paediatric Haematology,
King's College Hospital, Denmark Hill, London, UK

Fourth Edition

WILEY Blackwell

This fourth edition first published 2025

Edition History
John Wiley & Sons Ltd (3e, 2020); Blackwell Publishing Ltd. (2e, 2006)

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, New Era House, 8 Oldlands Way, Bognor Regis, West Sussex, PO22 9NQ

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
The contents of this work are intended to further general scientific research, understanding, and discussion only and are not intended and should not be relied upon as recommending or promoting scientific method, diagnosis, or treatment by physicians for any particular patient. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of medicines, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each medicine, equipment, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data
Names: Bain, Barbara J., author. | Rees, D. C. (David C.), author.
Title: Haemoglobinopathy diagnosis / Barbara J. Bain, David C. Rees.
Description: Fourth edition. | Hoboken : Wiley-Blackwell, 2025. |
Includes bibliographical references and index.
Identifiers: LCCN 2024034634 (print) | LCCN 2024034635 (ebook) | ISBN 9781394265350 (hardback) | ISBN 9781394265374 (adobe pdf) | ISBN 9781394265367 (epub)
Subjects: MESH: Hemoglobinopathies–diagnosis | Hematologic Tests–methods
Classification: LCC RC641.7.H35 (print) | LCC RC641.7.H35 (ebook) | NLM WH 190 | DDC 616.1/51075–dc23/eng/20241010
LC record available at https://lccn.loc.gov/2024034634
LC ebook record available at https://lccn.loc.gov/2024034635

Cover Design: Wiley
Cover Images: © Hector Roqueta Rivero/Getty Images, Courtesy of Barbara J. Bain

Set in 9/12pt Palatino by Straive, Pondicherry, India

Contents

Preface

This book is dedicated to the past and present scientific staff of the haematology departments of Princess Alexandra Hospital, Brisbane, Australia and St Mary's Hospital, Paddington, London (subsequently Imperial College Healthcare NHS Trust). It was the former group who first awakened Barbara Bain's interest in this field. They also suggested to her that there was a need for a practical book on the laboratory diagnosis of haemoglobinopathies and that she might be the person to write it. The second group, and in particular Lorry Phelan, for over four decades shared her pleasure in solving diagnostic problems and, at the same time, providing an accurate, clinically relevant diagnostic service; subsequent staff members did the same. It is with much pleasure that we dedicate this new edition to these colleagues and friends.

We should also like to acknowledge many other colleagues throughout the world who have helped in diverse ways, including those who have contributed images. They are individually acknowledged in the figure legends.

Barbara J. Bain and David C. Rees
London, 2024

Abbreviations and glossary

α the Greek letter alpha

α chain the α globin chain, which is required for 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, *HBA1* and *HBA2*, that encode α globin

α thalassaemia a group of thalassaemias characterised by absent or reduced α globin chain synthesis, usually resulting from deletion of one or more of the α globin genes; less often it results from altered structure of an α gene or mutation of the locus control gene, *LCRA*, 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 but not absent translation of α globin 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, *HBB*, that encodes β globin

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

β thalassaemia intermedia a β thalassaemia with significant clinicopathological abnormalities but not dependent on transfusion for survival; also known as non-transfusion-dependent thalassaemia

β thalassaemia major a β thalassaemia with significant clinicopathological abnormalities, requiring transfusion to sustain life

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, *HBG1* and *HBG2*, 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, *HBD*, that encodes δ globin

δ thalassaemia reduced or absent synthesis of δ globin and therefore of haemoglobin A_2

ε the Greek letter epsilon

ε chain the ε globin chain, which is synthesised 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, *HBE1*, that encodes ε globin chain

ψ the Greek letter psi, used to indicate a pseudogene

ζ the Greek letter zeta

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

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

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
AHSP α haemoglobin stabilising protein (AHSP)
AIDS acquired immune deficiency syndrome
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 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 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, at 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-methyl guanosine cap, added to RNA molecule during processing
capillary electrophoresis electrophoresis within a capillary tube
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-HLPC cation-exchange high performance liquid chromatography; *see* HPLC
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)
CNV copy number variant
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 atoms 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
COVID-19 corona virus disease 2019
CT computed tomography
CV coefficient of variation
DCIP test a screening test for haemoglobin E using 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 synthesised by the cell; the DNA molecule is a double helix of two complementary intertwined polynucleotides
EDTA ethylene diamine tetra-acetic acid
eIF2 erythroid initiation factor 2

EKLF erythroid Krüppel-like factor

electrophoresis separation of charged suspended particles such as proteins by application to a membrane or gel or within a capillary 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 or membrane

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 messenger RNA; most genes are composed of exons and non-translated introns

FAB classification French–American–British classification (of acute leukaemia)

FBC full blood count

Fe iron

Fe^{++}, **Fe^{2+}** ferrous or bivalent iron

Fe^{+++}, **Fe^{3+}** ferric or trivalent iron

fetal haemoglobin *see* haemoglobin F

G6PD glucose-6-phosphate dehydrogenase

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

GATA1 an erythroid-specific transcription factor

GDP guanosine diphosphate

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, for a minority of genes, 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 (c.f. 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

GTP guanosine triphosphate

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 almost all adults and, as an even lower proportion of total haemoglobin, in neonates and infants, having two α chains and two δ chains

haemoglobin A_2' a haemoglobin A_2 variant, also known as haemoglobin B_2

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 alpha 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, F or Gower 2

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 that is synthesised 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 percentage saturation of haemoglobin against 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, having

two α chains and two γ chains; also 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 ζ chains and two ε 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 A and J on electrophoresis at alkaline pH

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

haemoglobin M a variant haemoglobin that oxidises readily to methaemoglobin

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

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

haemoglobinopathy an inherited disorder resulting from synthesis of a structurally abnormal haemoglobin; the term can also be used to encompass, in addition, 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 abnormal embryonic haemoglobin, having two ζ chains and two β chains, present in some severe thalassaemia syndromes

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

Hb haemoglobin concentration

Hct haematocrit

HDAC1 histone deacetylase 1

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

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

homologous being equivalent or similar to another

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 or X chromosome gene

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

HPFH hereditary persistence of fetal haemoglobin

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

HRI haem-regulated inhibitor
HS1, HS2, HS3, HS4 hypersensitive sites 1, 2, 3, and 4, upstream of the β globin gene cluster
HS −40 an upstream enhancer of α globin gene transcription, part of *LCRA*
HVR hypervariable region
ICSH International Council for Standardization in Haematology
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 that 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
intervening sequence an intron
intron a sequence of DNA in a gene that is not represented in processed messenger RNA 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, intron
kb kilobase, a unit for measuring the length of DNA; one kilobase is 1000 nucleotide base pairs
kD kilodalton, a unit for measuring molecular weight; one kilodalton is 1000 daltons
KLF1 Krüppel-like factor 1
LCR locus control region, a DNA sequence upstream of genes of the α or β globin cluster that enhances transcription of the genes of the cluster, *LCRA* and *LCRB* control the α and β gene clusters respectively
LCRA locus control region alpha
LCRB locus control region beta
LDH lactate dehydrogenase
MCH mean cell haemoglobin
MCHC mean cell haemoglobin concentration
MCV mean cell volume
MDS myelodysplastic syndrome
methaemoglobin oxidised haemoglobin, which does not function in oxygen transport
MGG May–Grünwald–Giemsa (stain)
mis-sense mutation a mutation that leads to the encoding of a different amino acid
MPLA multiple ligation-dependent probe amplification
MRI magnetic resonance imaging
mRNA messenger RNA, 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 synthesis of proteins
MS mass spectrometry, electrospray ionisation mass spectrometry, a method for determining the mass and the charge of a molecule
NO nitric oxide
nonsense mutation a mutation that leads to no amino acid being encoded that therefore functions as a STOP or termination codon, leading to synthesis of a truncated polypeptide chain
non-transfusion-dependent thalassaemia a thalassaemia with significant clinicopathological abnormalities but not dependent on transfusion for survival, also known as β thalassaemia intermedia
NRBC nucleated red blood cell/cells
NTDT non-transfusion-dependent thalassaemia
O_2 oxygen
ORF open reading frame
oxyhaemoglobin haemoglobin combined with O_2
P_{50} PO_2 at which haemoglobin is half saturated
***P*aO_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 (stain)
PCR polymerase chain reaction, a method of making multiple copies of a DNA sequence
PCV packed cell volume, haematocrit
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 (c.f. genotype)

PO_2 partial pressure of oxygen

polymorphism the occurrence of a variant form of a gene in a significant proportion (at least 1%) of a population

promoter a sequence of DNA at the 5′ end of a gene that 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, a measure of anisocytosis

restriction endonuclease an enzyme that recognises 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 demonstration of a specific gene that removes or creates a specific cleavage site

ribosome a cytoplasmic structure on which proteins are translated from messenger RNA; 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

SARS-CoV-2 severe acute respiratory syndrome-corona virus-2

sickle cell an erythrocyte that has become sickle or crescent shaped as a result of polymerisation 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 (preferred definition but sometimes used as a synonym for sickle cell anaemia)

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

SNP single nucleotide polymorphism

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

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

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

thalassaemia intermedia a genetically heterogeneous thalassaemic condition that is moderately severe but nevertheless does not require regular blood transfusions to sustain life; usually refers to β thalassaemia intermedia

thalassaemia major thalassaemia that is incompatible with more than a short survival in the absence of blood transfusion; usually refers to β thalassaemia major

thalassaemia minor an asymptomatic thalassaemic condition, attributable to β thalassaemia heterozygosity or to deletion of one or two of the four α genes; usually referred to as thalassaemia trait

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 a membranous sac attached to 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 has a molecular weight of 64–64.5 kDa and is composed of two dissimilar pairs of polypeptide chains, each of which encloses an iron-containing porphyrin designated haem (Fig. 1.1). Haem is essential for oxygen transport while globin serves to protect haem from oxidation, renders the molecule soluble and permits variation in oxygen affinity. The structure of the haemoglobin molecule produces an internal environment of hydrophobic radicals, which protects the iron of haem from water and thus from oxidation. External radicals are hydrophilic and thus render the haemoglobin molecule soluble. Both haem and globin are subject to modifications. The iron of haemoglobin is normally in the ferrous form (Fe^{2+}). Haem is able to combine reversibly with oxygen so that haemoglobin can function as an oxygen-transporting protein. Oxidation of iron to the ferric form (Fe^{3+}) is a less readily reversible reaction, converting haem to haematin and haemoglobin to methaemoglobin, a form of haemoglobin that cannot transport oxygen. Auto-oxidation of haemoglobin to methaemoglobin is a normal process. About 3% of haemoglobin undergoes this process each day with about 1% (0.4–1% in one study) of haemoglobin being methaemoglobin [1, 2]. Methaemoglobin is converted back to haemoglobin mainly by the action of NADH-cytochrome b5-methaemoglobin reductase.

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

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

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

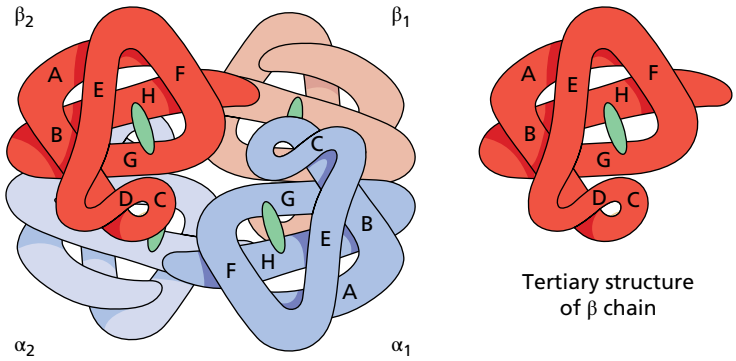


Fig. 1.1 Diagrammatic representation of the tertiary structure of a haemoglobin monomer (a β globin chain containing a haem group) and the quaternary structure of haemoglobin; upper case letters indicate homologous α helixes.

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

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

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

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

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

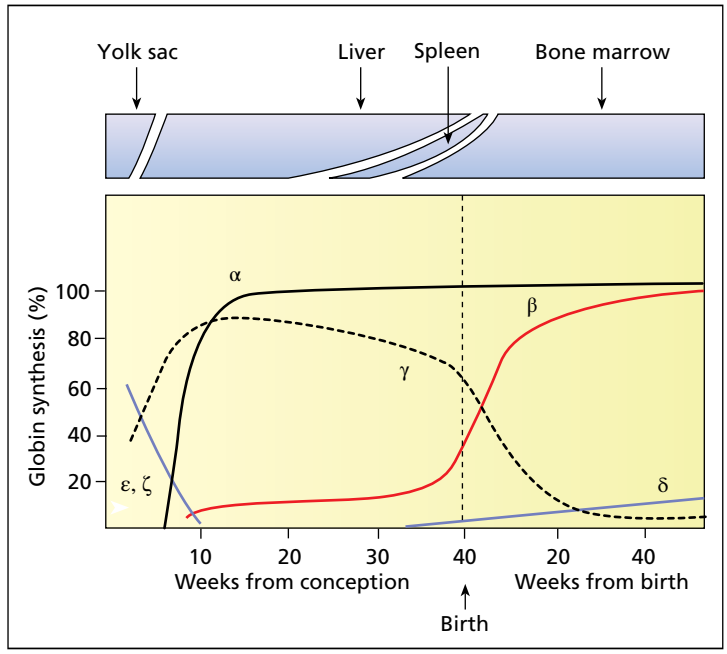


Fig. 1.2 Diagrammatic representation of the sites and rates of synthesis of different globin chains in the embryonic and fetal periods and during infancy.

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

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

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

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

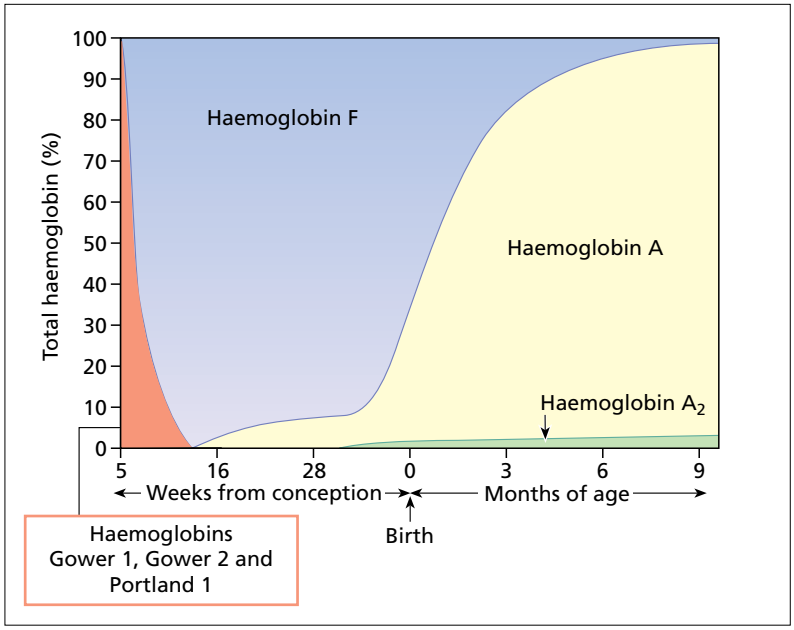


Fig. 1.3 Diagrammatic representation of the average percentages of various haemoglobins present in the embryonic and fetal periods and during infancy.

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

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

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

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

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

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

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

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

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

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

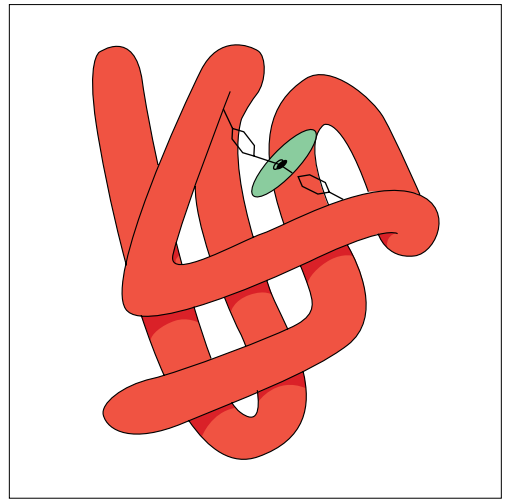


Fig. 1.4 Diagrammatic representation of a haemoglobin molecule with a haem group within the haem pocket, showing the relationship of the haem to two histidine residues of the globin chain, designated proximal and distal histidines; haem is bound to the proximal histidine while O_2 is bound to haem and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

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

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

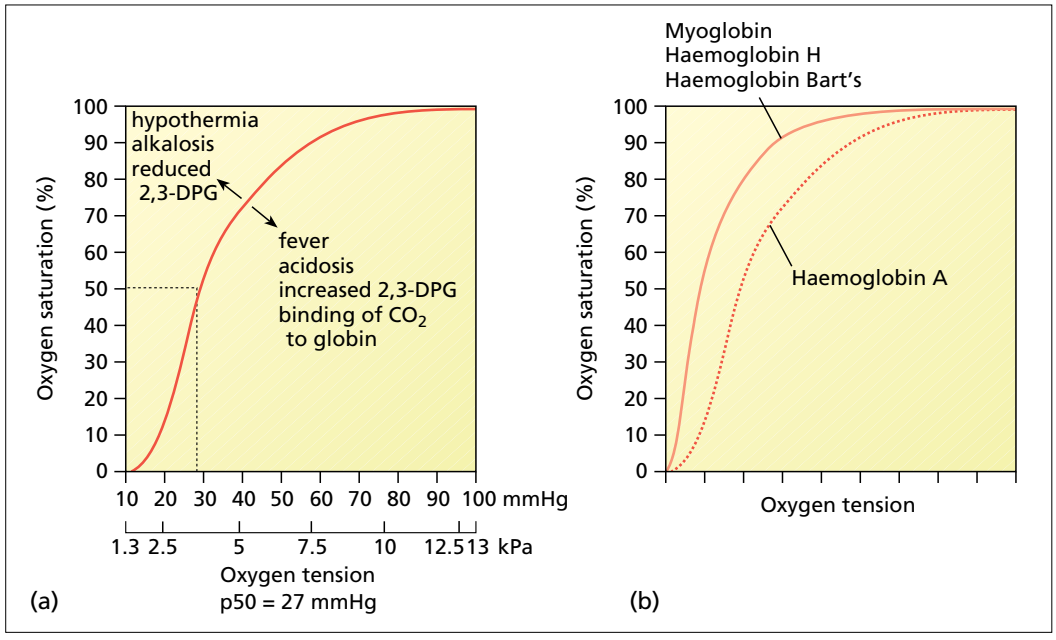


Fig. 1.5 Oxygen dissociation curve: (a) normal oxygen dissociation curve indicating the effects of alteration of pH, body temperature and 2,3-diphosphoglycerate (2,3-DPG) concentration on the oxygen affinity of haemoglobin; (b) a comparison of the hyperbolic oxygen dissociation curve characteristic of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

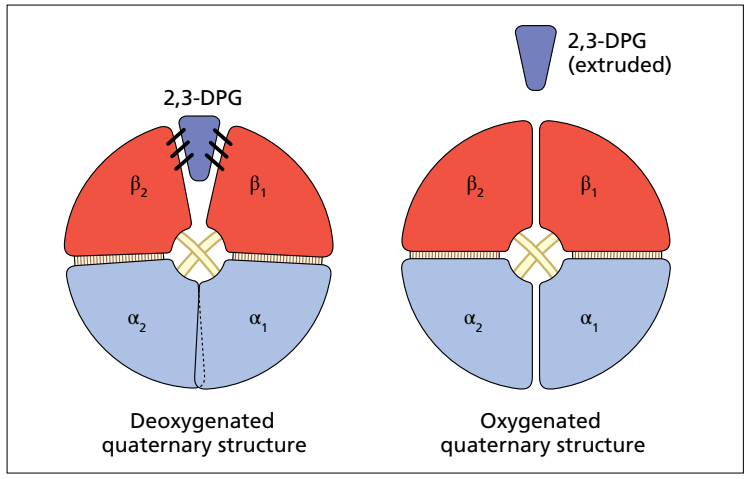


Fig. 1.6 Diagrammatic representation of the effect of oxygenation and deoxygenation on the quaternary structure of haemoglobin. The haemoglobin dimers ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) are stable, with the dimeric bonds between the α and the β chain having 34 contacts in both the deoxygenated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

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

Genetics of haemoglobin synthesis

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

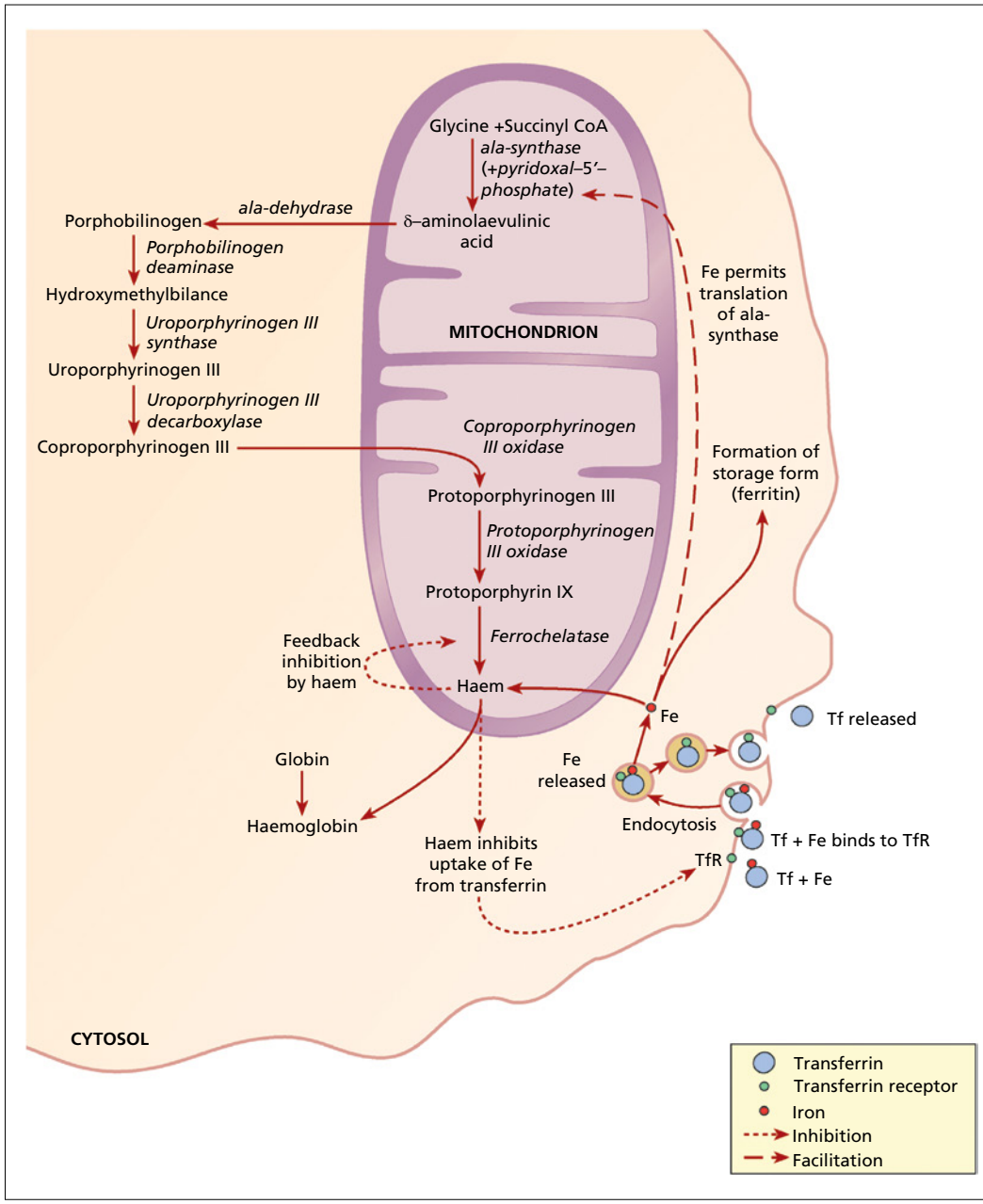


Fig. 1.7 Diagrammatic representation of haem synthesis. Tf, transferrin; TfR, transferrin receptor.

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

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

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

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

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

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

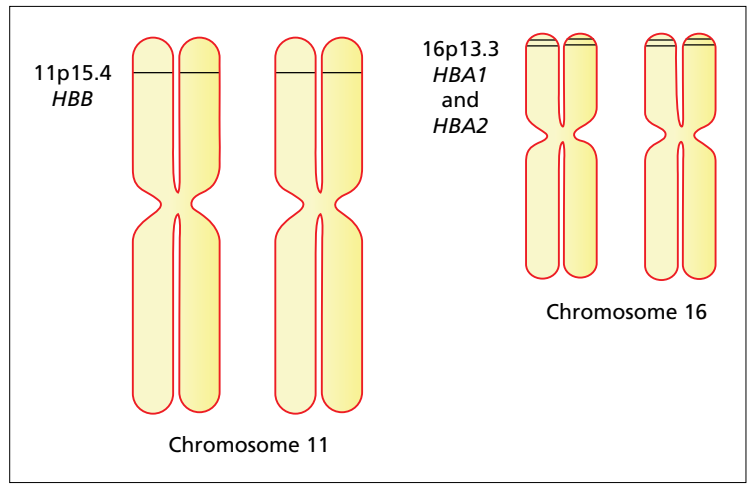


Fig. 1.8 Diagram of chromosomes 11 and 16 showing the positions of the β and α globin gene clusters.

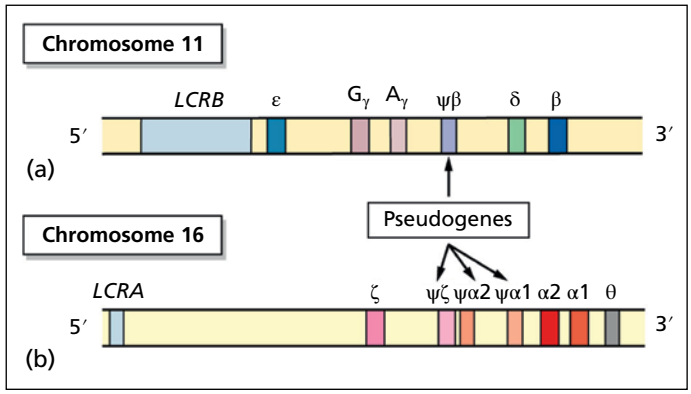


Fig. 1.9 Diagrammatic representation of the α and β globin gene clusters on chromosome 11 (a) and chromosome 16 (b).

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

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

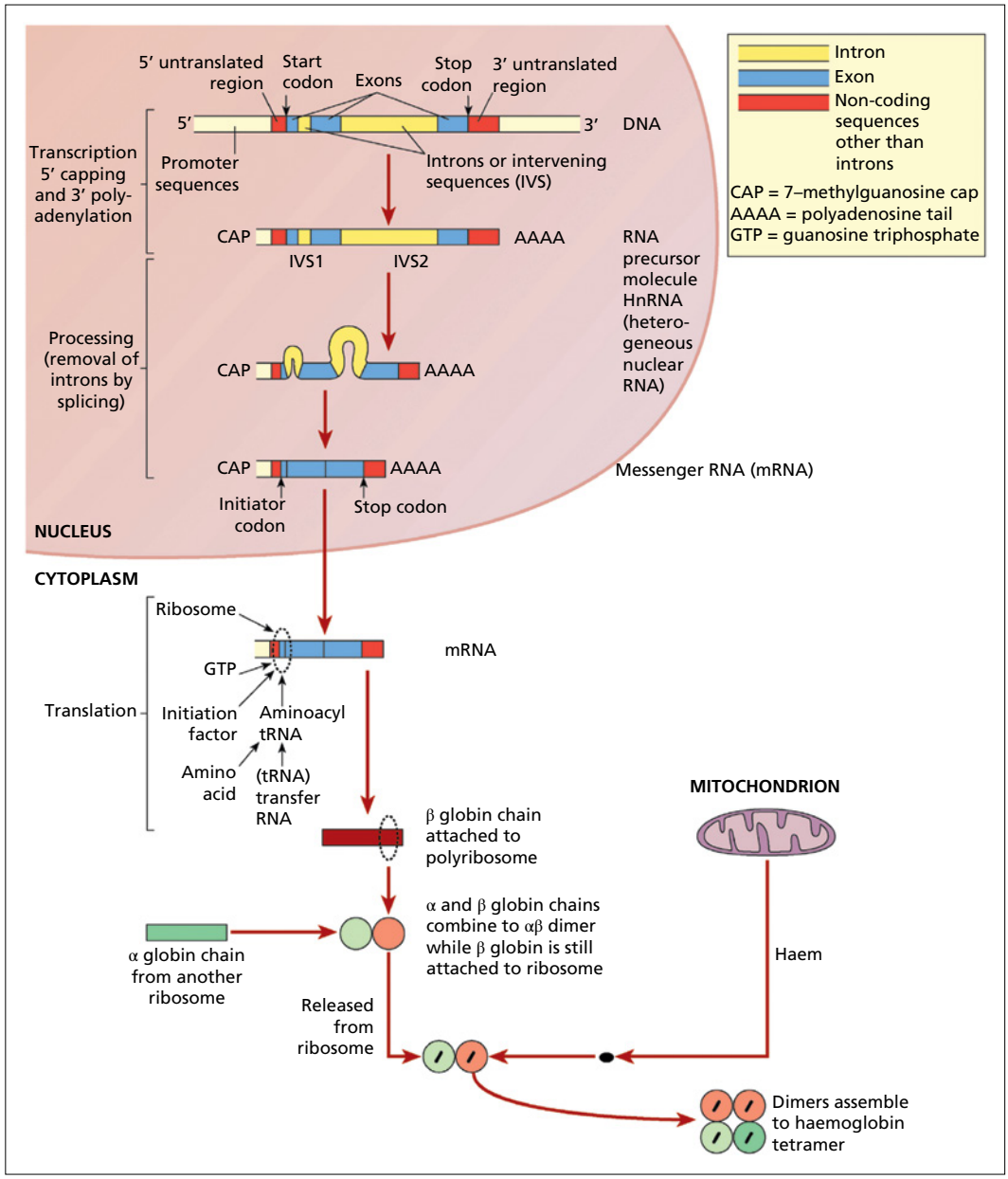


Fig. 1.10 Diagrammatic representation of ribonucleic acid (RNA) synthesis and processing and β globin chain synthesis. Although processes are shown sequentially, capping starts soon after transcription has started and therefore contemporaneously with transcription whereas polyadenylation necessarily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

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

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

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

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

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

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