OXFORD

HEMOGLOBIN

Insights into Protein Structure, Function, and Evolution

JAY F. STORZ

Hemoglobin

Hemoglobin

Insights into Protein Structure, Function, and Evolution

Jay F. Storz
Professor, School of Biological Sciences,
University of Nebraska

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2019
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2018950707

ISBN 978–0–19–881068–1 (hbk.)
ISBN 978–0–19–881069–8 (pbk.)

DOI: 10.1093/oso/9780198810681.001.0001

Printed in Great Britain by
Bell & Bain Ltd., Glasgow

For my kids, Jessie and Cody

Preface

> "More may have been written about hemoglobin than about any other molecule. Physicists, crystallographers, chemists of all kinds, zoologists, physiologists and geneticists, pathologists, and hematologists have all contributed to a vast literature. In the erratic ways that scientific research shares with other human endeavors, the multifarious work of that great throng has provided us with an enormous store of knowledge from which one can extract data on subjects as diverse as the quantum chemistry of iron and the buoyancy of fish."
>
> **—Perutz (2001)**

In the chapters to follow, we will explore questions about protein structure and function, biochemical adaptation, and molecular evolution by focusing on lessons learned from research on a single, paradigmatic protein. There are several reasons to focus specifically on hemoglobin (Hb) and other members of the globin superfamily; one is historical importance. During the last half century, Hb has played a starring role in research efforts to understand relationships between protein structure and function, and in efforts to identify the molecular underpinnings of physiological adaptation and pathophysiology. Hb and its cousin, myoglobin (Mb), were the first proteins to have their crystal structures solved, Hb serves as a paradigm for understanding principles of allosteric regulatory control, and clinical research on mutant Hbs ushered in the modern era of molecular medicine.

As stated by Dickerson and Geis (1983): "with this one family of macromolecules [Hb and related globin proteins] one can illustrate nearly every important feature of protein structure, function, and evolution: principles of amino acid sequence and protein folding, a mechanism of activity that resembles that found in enzymes although not itself catalytic, specificity in the recognition and binding of large and small molecules, subunit motion and allosteric control in regulating activity, gene structure and genetic control, and the effects of point mutations on molecular behavior." Dickerson and Geis also emphasized the merits of familiarity: "People for whom alpha-ketoglutarate has no charms immediately recognize hemoglobin as the essential constituent of blood, without which human life would be impossible."

The other rationale for focusing specifically on Hb is based on the time-tested scientific practice of using a model system to extrapolate general principles. As foundational knowledge accumulates, the value of the model continues to increase. For example, solving the crystal structure of a protein can provide insights into the stereochemical basis of observed functional properties. These insights can then suggest new hypotheses about structure-function relationships which, in turn, motivate further experimental work. Research on Hb structure, function, and evolution illustrates how a well-chosen model system can enhance our investigative acuity and bring key questions into focus.

This book is aimed at an interdisciplinary audience, including evolutionary biologists with an interest in how mechanistic studies of protein function can be used to address fundamental questions about evolution, and biochemists and physiologists with an interest in how evolutionary approaches can broaden and enrich their field of study. Most previous book-length treatments of Hb structure and

function have devoted much space to discussions of Hb disorders such as sickle-cell anemia and various forms of thalassemia (Dickerson and Geis, 1983, Bunn and Forget, 1986). The volume by Steinberg et al. (2009) provides an especially authoritative and complete compendium of information about Hb-related diseases. The present volume has a different aim, so I have not devoted much space to the discussion of pathophysiology except in cases where understanding the etiologies of particular Hb disorders helps illustrate a broader point. Relative to previous books about Hb structure and function, I have sharpened the focus on conceptual issues of relevance to questions about biochemical adaptation and mechanisms of protein evolution.

To lay the foundation, Chapter 1 reviews basic principles of protein structure—the nature of proteins as polymers of amino acids, the variety of amino acids, and the way in which the physicochemical properties of amino acid side chains influence the folding of a polymer into a three-dimensional protein with specific functional properties. Chapter 2 then provides an overview of Hb function and its physiological role in respiratory gas transport. Much of the chapter is devoted to explaining the physiological significance of cooperative O_2 binding by Hb and therefore provides a point of departure for Chapter 3, which provides a brief overview of allosteric theory. Chapter 4 provides an overview of Hb structure and explains the mechanistic basis of allosteric effects. Chapter 5 provides an overview of the evolutionary history of the globin gene superfamily and places the evolution of vertebrate-specific globins in phylogenetic context. The remaining chapters explore the physiological significance of gene duplication and Hb isoform differentiation (Chapter 6), the evolution of novel Hb functions and physiological innovation (Chapter 7), and mechanisms of biochemical adaptation to environmental hypoxia (Chapter 8). Finally, Chapter 9 discusses conceptual issues in protein evolution and provides a synthesis of lessons learned from studies of Hb.

During the preparation of this book, I have been fortunate to receive helpful suggestions from friends and colleagues over the world: Andrea Bellelli (Sapienza University of Rome, Italy), Michael Berenbrink (University of Liverpool, UK), Colin J. Brauner (University of British Columbia, Canada), Thorsten Burmester (University of Hamburg, Germany), Kevin L. Campbell (University of Manitoba, Canada), Angela Fago (Aarhus University, Denmark), Chien Ho (Carnegie Mellon University), Federico G. Hoffmann (Mississippi State University), David Hoogewijs (University of Duisburg-Essen, Germany), Frank B. Jensen (University of Southern Denmark, Denmark), Amit Kumar (State University of New York-Buffalo), Hideaki Moriyama (University of Nebraska), Chandrasekhar Natarajan (University of Nebraska), John S. Olson (Rice University), Juan C. Opazo (Universidad Austral de Chile, Chile), William E. Royer (University of Massachusetts Medical School), Graham R. Scott (McMaster University, Canada), Anthony V. Signore (University of Nebraska), Jeremy R. H. Tame (Yokohama City University, Japan), Tobias Wang (Aarhus University, Denmark), Roy E. Weber (Aarhus University, Denmark), and Mark A. Wilson (University of Nebraska). I am also very grateful to F. G. Hoffmann, A. Kumar, and C. Natarajan for their help with figures.

I thank Angela Fago and Roy Weber for their part in maintaining a fun and productive transatlantic collaboration built on strong friendship. Our work together inspired me to write this book.

Finally, I thank my wife, Eileen, our kids, Jessie and Cody, and our dog, Gwydion, for helping to ensure that time spent writing was well-balanced with other pursuits!

Jay F. Storz
Lincoln, Nebraska

References

Bunn, H. F. and Forget, B. G. (1986). *Hemoglobin: Molecular, Genetic and Clinical Aspects*, Philadelphia, PA, W. B. Saunders Company.

Dickerson, R. E. and Geis, I. (1983). *Hemoglobin: Structure, Function, Evolution, and Pathology*, Menlo Park, CA, Benjamin/Cummings.

Perutz, M. F. (2001). Molecular anatomy and physiology of hemoglobin. *In:* Steinberg, M. H., Forget, B. G., Higgs, D. R., and Nagel, R. L. (eds.) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*, pp. 174–96. Cambridge, UK, Cambridge University Press.

Steinberg, M. H., Forget, B. G., Higgs, D. R., and Weatherhall, D. J. (2009). *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*, 2nd edition, Cambridge, UK, Cambridge Medicine.

Contents

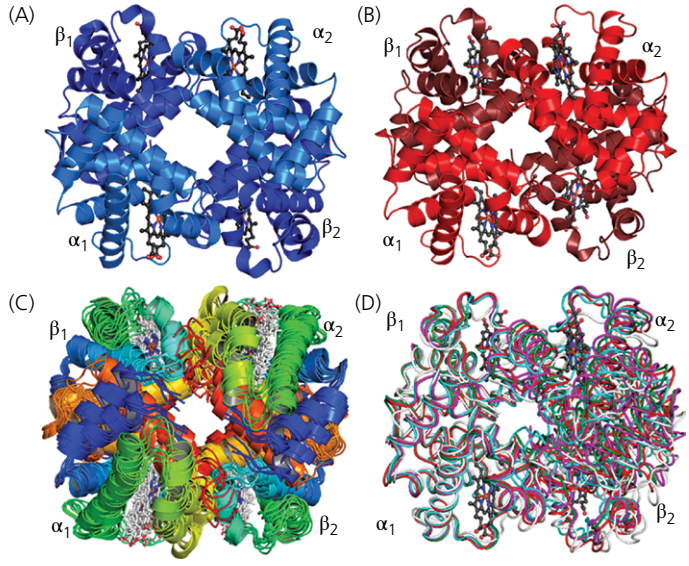

Plate 1. Structures of human Hb. (A) Crystal structure of deoxy Hb (PDB code 2DN2); (B) Crystal structure of HbCO (2DN3); (C) Ten lowest energy solution structures of HbCO obtained by NMR spectroscopy (2M6Z); (D) Superimposition of different R-type crystal structures of HbCO (2DN3, red; 1BBB, magenta; 1MKO, green; and 1YZI, cyan) with the average solution structure obtained by means of NMR (light gray). Structures were aligned according to the $\alpha_1\beta_1$ dimer.

Reproduced from Yuan et al. (2015), with permission from the American Chemical Society. See Chapter 2, section 2.1.

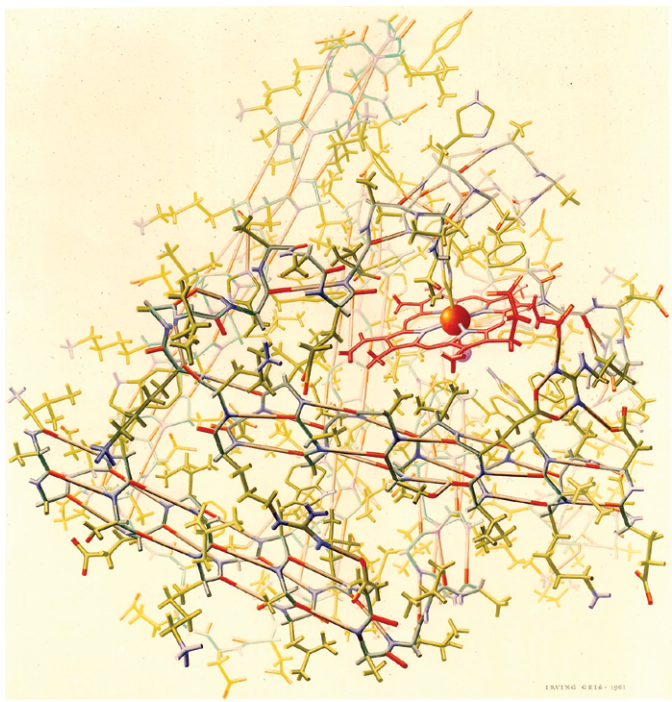

Plate 2. Skeletal drawing of the myoglobin protein by the pioneering scientific illustrator, Irving Geis (1908–1997). This illustration, which appeared in a 1961 issue of *Scientific American*, was the first full side-chain drawing of any protein molecule. Today, such renderings are created by computer.

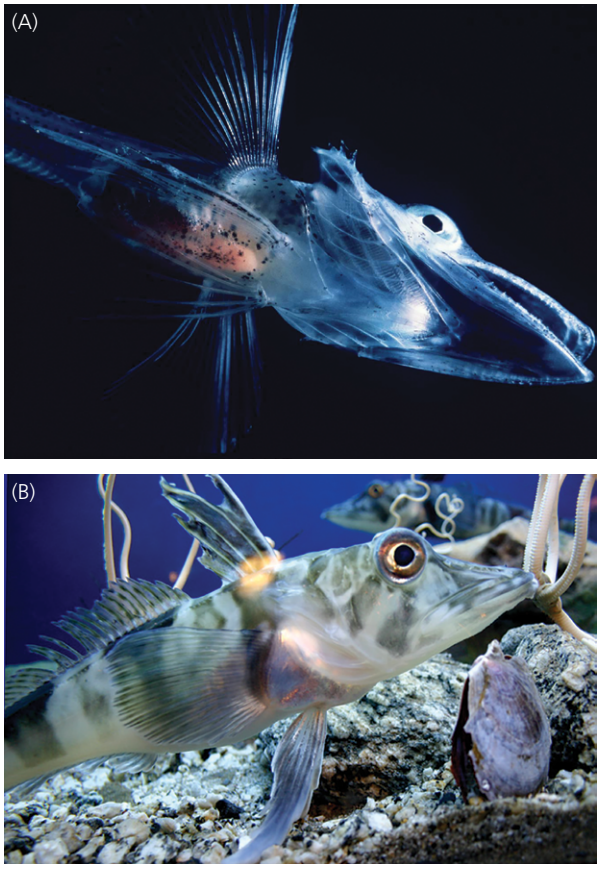

Plate 3. Icefish in the family Channichthyidae inhabit the freezing, ice-laden waters surrounding the continental shelf of Antarctica. These physiologically enigmatic fish do not express Hb and therefore have colorless blood. (A) Icefish larva (Uwe Kils, Wikimedia Commons) and (B) adult crocodile icefish, *Chionodraco hamatus* (Marrabbio2, Wikimedia Commons). See Chapter 2, section 2.6.

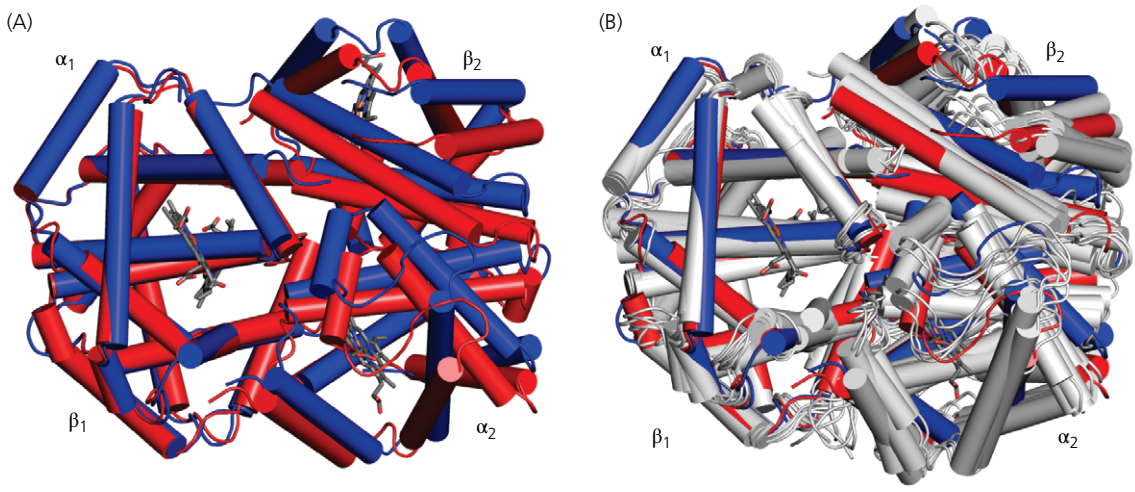

Plate 4. Hb allostery was originally viewed in terms of a rigid two-structure model (A), but is now interpreted in terms of a dynamic ensemble of structures (B). Panel A shows superimposed structures of human Hb in the T-state (unliganded conformation, PDB code 4HHB, blue) and in the R-state (liganded, conformation, PDB code 2DN3, red). Panel B shows the ten lowest energy solution structures of HbCO obtained via NMR (PDB code 2M6Z, gray) superimposed with the canonical T and R structures determined by X-ray crystallography. Structures were aligned according to the $\alpha_1\beta_1$ dimer.

Reproduced from Yuan et al. (2015), with permission from the American Chemical Society. See Chapter 4, section 4.3.

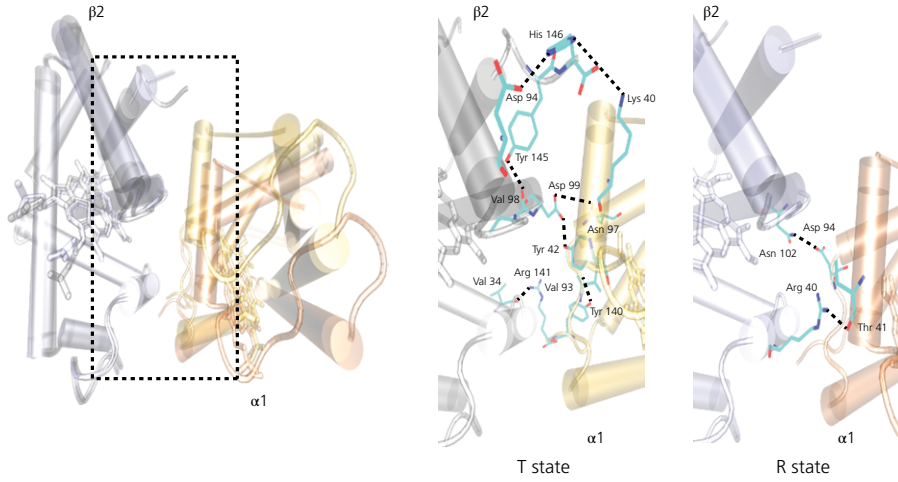

Plate 5. Network of atomic contacts at interdimer ($\alpha_1\beta_2/\alpha_2\beta_1$) interfaces involving the C and G helices and FG corner. These interdimer "sliding" contacts involve α_1C-β_2FG and α_1FG-β_2C, and the minor contacts involve α_1C-β_2C and α_1FG-β_2FG. The complete set of subunit contacts are detailed in Table 4.1. During the oxygenation-linked transition from the T-state to the R-state, some interactions at the $\alpha_1\beta_2/\alpha_2\beta_1$ interface are broken and new interactions are formed. As stated by Perutz (1978), this intersubunit interface "acts as a snap-action switch, with two alternative stable positions, each braced by a different set of hydrogen bonds." Intermediate positions of the interface are sterically prevented, consistent with the all-or-nothing conformational transition envisioned by the MWC model. See Chapter 4, section 4.3.

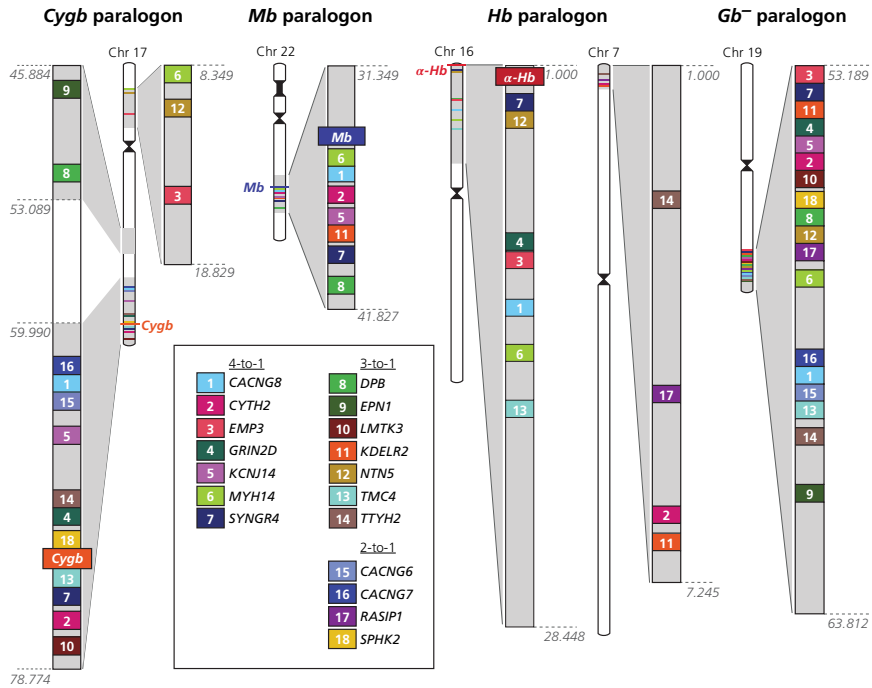

Plate 7. Fourfold pattern of conserved macrosynteny between the four globin-defined paralogons in the human genome (including the *Gb*⁻ paralogon) and "linkage group 15" of the reconstructed proto-karyotype of the chordate common ancestor (shaded regions). This pattern of conserved macrosynteny demonstrates that the *Cygb*, *Mb*, *Hb*, and *Gb*⁻ paralogons trace their duplicative origins to the same proto-chromosome of the chordate common ancestor and provides conclusive evidence that each of the four paralogons are products of a genome quadruplication in the stem lineage of vertebrates. Shared gene duplicates that map to secondarily translocated segments of the *Mb* paralogon (on chromosome 12) and the *Hb* paralogon (on chromosomes 7 and 17) are not pictured.

Reproduced from Hoffmann et al. (2012b). See Chapter 5, section 5.8.

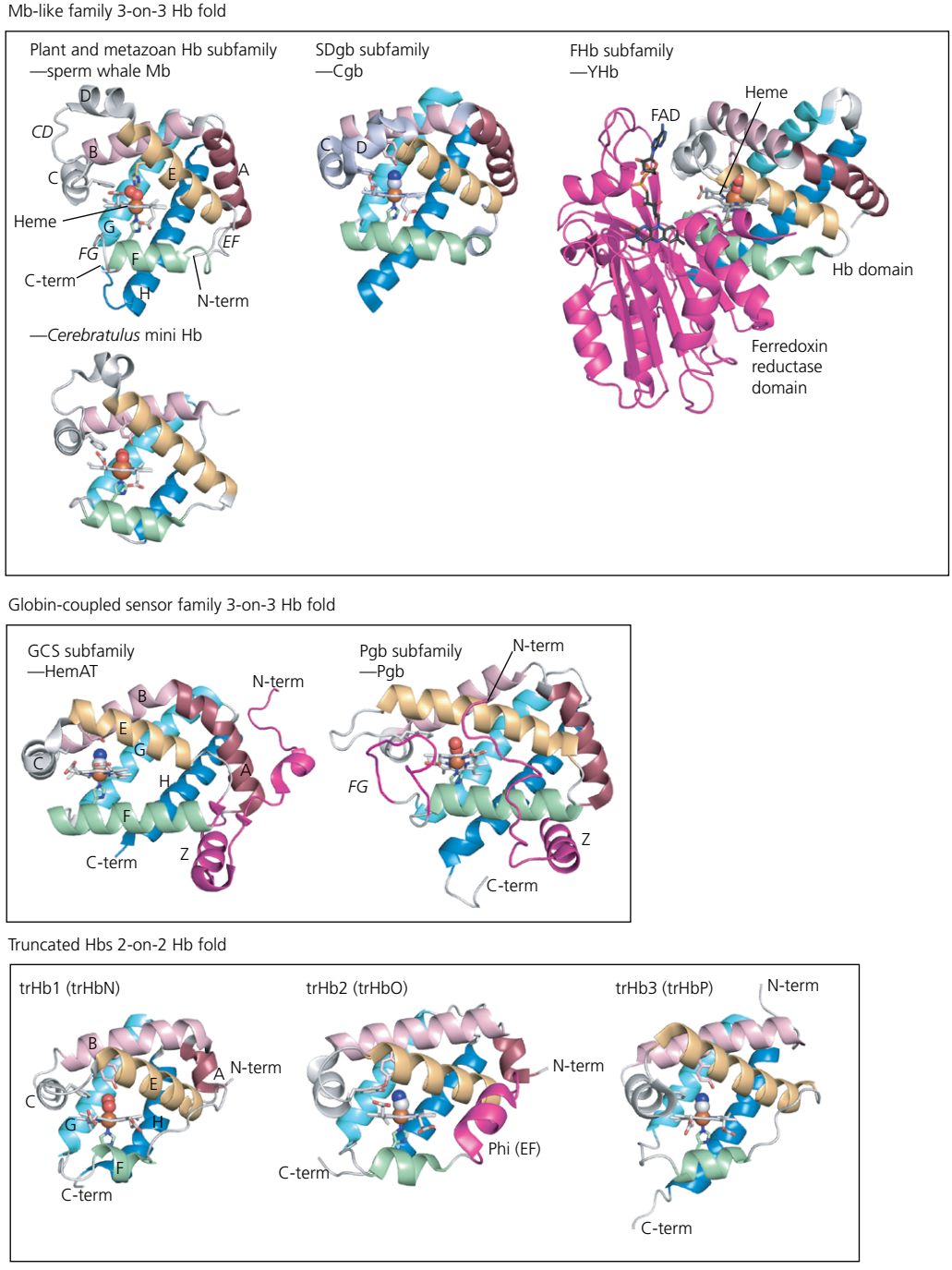

Plate 6. Conserved and variable features of globin tertiary structure across different domains of life. The depicted tertiary structures are representative of major lineages comprising the globin protein superfamily. Most globins from the plant and metazoan "Hb subfamily" are highly similar to sperm whale Mb (pdb 2mgm), the exemplar of the Mb fold. Variations on the theme include the *Cerebratulus lacteus* mini Hb (1kr7). Other families and subfamilies are represented by Cgb from *Campylobacter jejuni* (2wy4), YHb from *Saccharomyces cerevisiae* (4g1v), HemAT from *Bacillus subtilis* (1or4), Pgb from *Methanosarcina acetivorans* (2veb), trHb1 (trHbN) from *Tetrahymena pyriformis* (3aq5), trHb2 (trHbO) from *Mycobacterium tuberculosis* (1ngk), and trHb3 (HbP) from *Campylobacter jejuni* (2ig3). Conserved α-helices that comprise the canonical 3-on-3 tertiary structure are color-coded as follows: A (red-brown), B (pink), E (yellow/tan), F (green), G (cyan), H (blue). Functionally important loops (CD, EF, FG) are also labeled. Conserved elements of the 2-on-2 globin fold are color-coded in the same fashion as the 3-on-3 globins. Additional elements of secondary structure that are present in Mb but not in some other globins are shown in gray. Additional structural elements that are unique to specific globin subfamilies are shown in magenta.

Modified from Gell (2017). See Chapter 5, section 5.3.

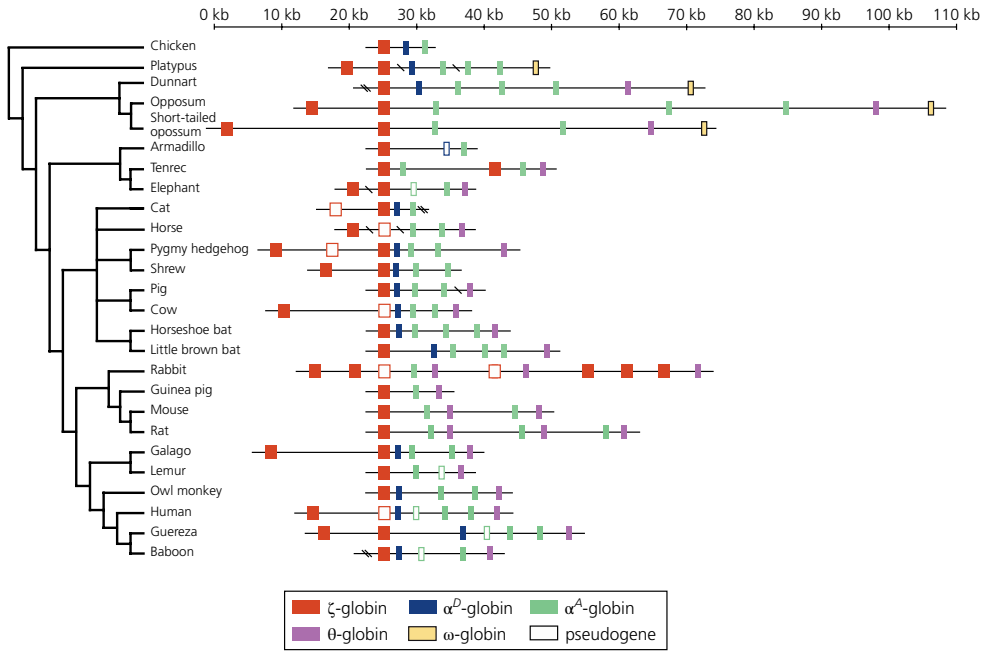

Plate 8. Variation in the size and membership composition of the α-globin gene family in mammals.

Reproduced from Hoffmann et al. (2008b).

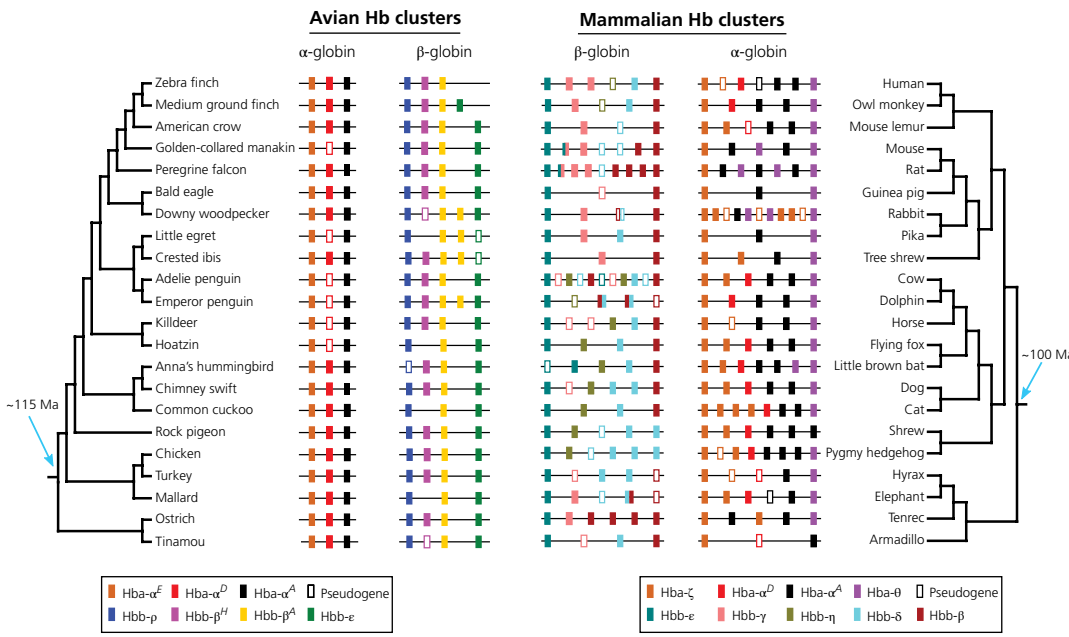

Plate 9. Dramatic differences in stability of the globin gene clusters in birds and mammals. Genes in the mammalian clusters have undergone a much higher rate of turnover. Shown are diagrammatic depictions of the chromosomal organization of the α- and β-type globin genes. In the case of the α-type globin genes, birds and mammals share orthologous copies of the α^D- and α^A-globin genes (α^D-globin is annotated as μ-globin in the human genome assembly). Likewise, the avian π-globin and the mammalian ζ-globin genes are 1:1 orthologs. In contrast, the genes in the avian and mammalian β-globin gene clusters are derived from independent duplications of one or more β-type globin genes that were inherited from the common ancestor of tetrapod vertebrates (Hoffmann et al., 2010).

Reproduced from Zhang et al. (2014). See Chapter 5, section 5.9.1.

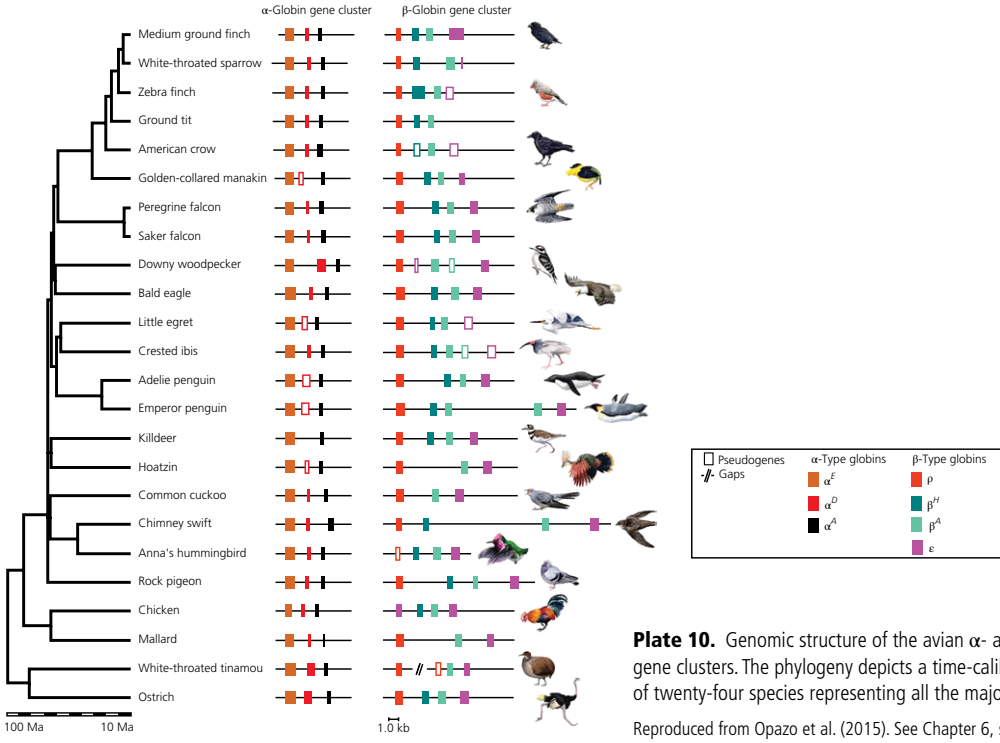

Plate 10. Genomic structure of the avian α- and β-type globin gene clusters. The phylogeny depicts a time-calibrated supertree of twenty-four species representing all the major avian lineages.

Reproduced from Opazo et al. (2015). See Chapter 6, section 6.6.1.

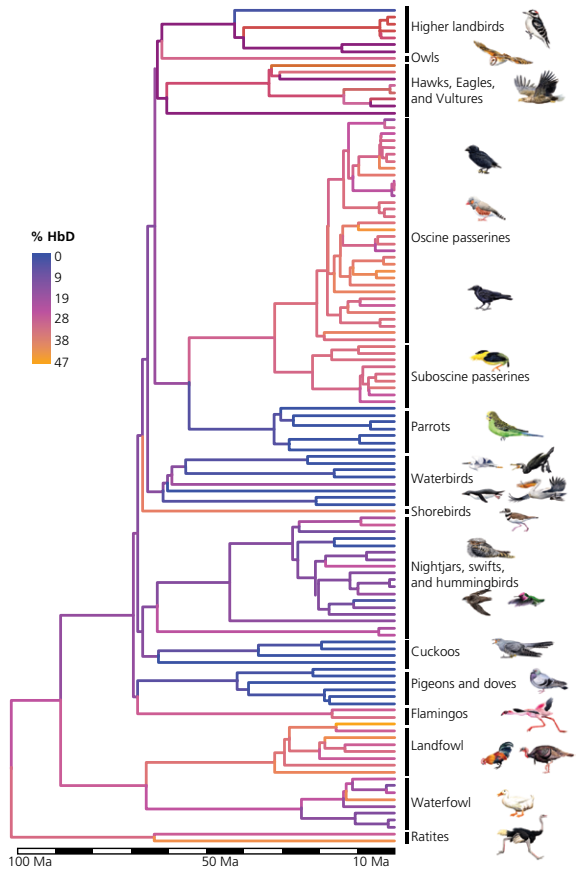

Plate 11. Inferred evolutionary changes in relative expression level of the HbD isoform (percentage of total Hb) during the diversification of birds. Expression data are based on experimental measures of protein abundance in the definitive red blood cells of 122 bird species (n = 1–30 individuals per species, 267 specimens in total). Terminal branches are color-coded according to the measured HbD expression level of each species, and internal branches are color-coded according to maximum-likelihood estimates of ancestral character states. Branch lengths are proportional to time.

Reproduced from Opazo et al. (2015). See Chapter 6, section 6.6.1.

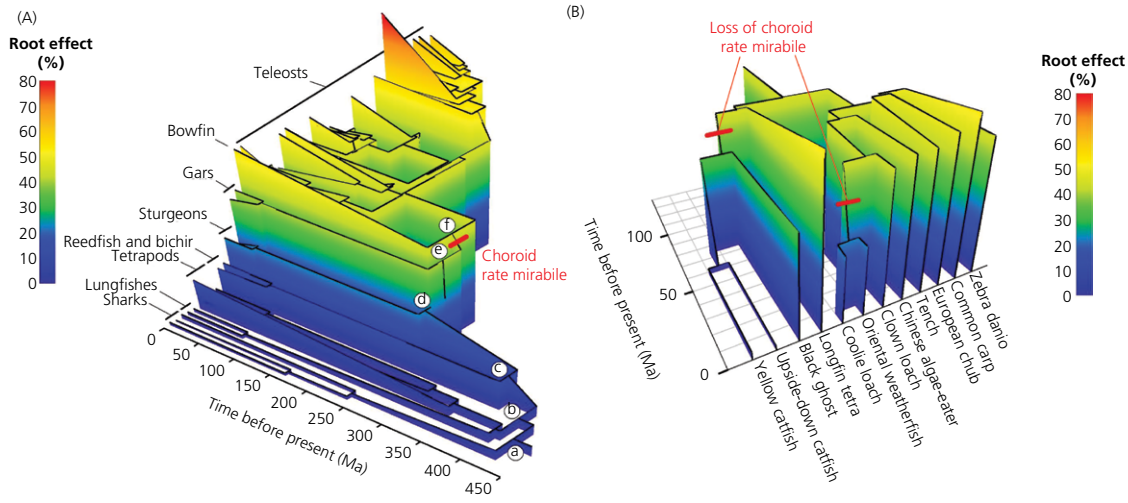

Plate 12. Evolution of the Root effect in jawed vertebrates. Ancestral values at internal nodes of the tree were estimated using linear parsimony (for methodological details, see Berenbrink et al. [2005]). (A) Rotation of the three-dimensional projection shows the gradual increase of the Root effect in early ray-finned fishes (nodes c–f) after divergence from the line leading to modern-day lobe-finned fishes (including tetrapods). The red hash mark denotes the origin of the choroid *rete mirabile* in the branch leading to the common ancestor of bowfin and teleosts. Note that the choroid *rete* originated only after the Root effect had increased to more than 40 percent. (B) Zoomed-in view of a clade of teleost fishes showing secondary reductions of the Root effect in Ostariophysi. Secondary reductions of the Root effect only occur in lineages in which the choroid *rete* has been lost (denoted by red hash marks).

Ma = million years.

Modified from Berenbrink (2007). See Chapter 7, section 7.2.1.

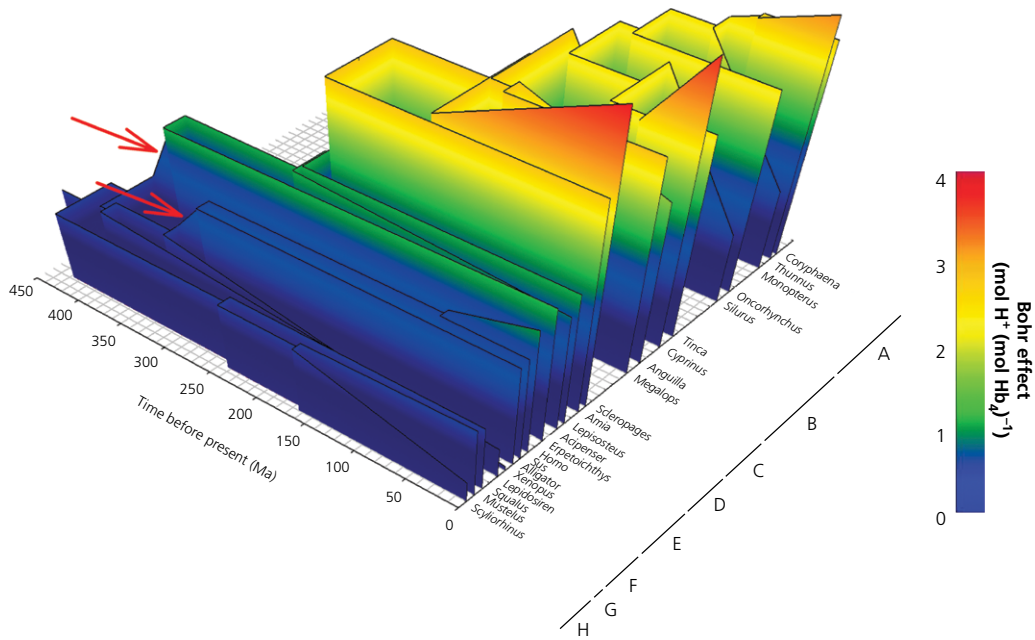

Plate 13. Evolution of the Bohr effect in jawed vertebrates. Values refer to the maximal alkaline Bohr effect of organic phosphate-free hemolysates in 0.1 M KCl at physiological temperature, determined by titration of hemolysates equilibrated under N_2 and O_2 atmospheres and calculated as the number of additional protons bound upon deoxygenation of Hb at constant pH. Ancestral values at internal nodes of the tree were estimated using linear parsimony (for methodological details, see Berenbrink et al. [2005]). Red arrows denote two independent increases of the Bohr effect in the stem lineage of tetrapods (F) and early ray-finned fishes (E).

Modified from Berenbrink et al. (2005). See Chapter 7, section 7.2.1.

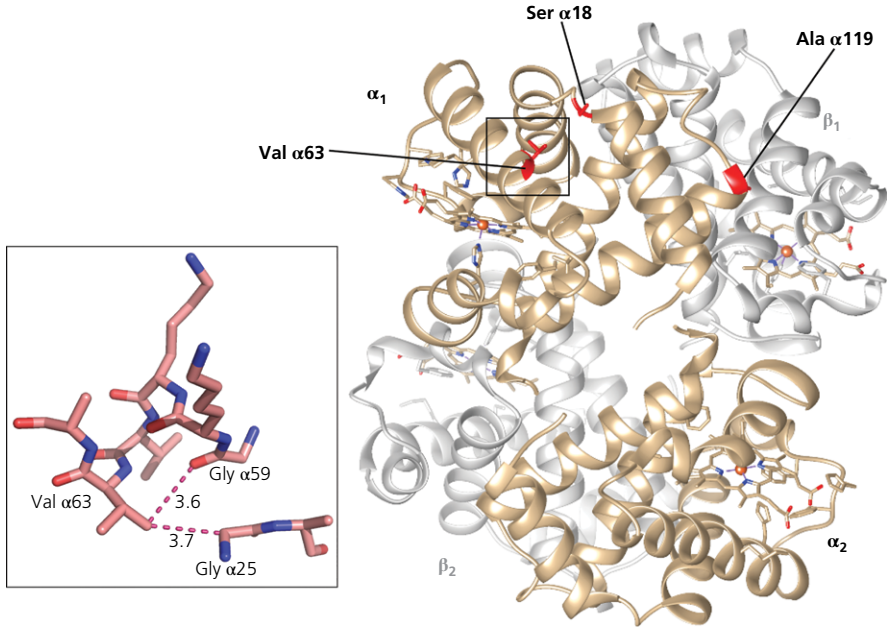

Plate 14. Structural model showing bar-headed goose Hb in the deoxy state (PDB1hv4), along with locations of each of the three amino substitutions that occurred in the bar-headed goose lineage after divergence from the common ancestor of other *Anser* species. The inset graphic shows the environment of the Val α63 residue. When valine replaces the ancestral alanine at this position, the larger volume of the side chain causes minor steric clashes with two neighboring glycine residues, Gly α25 and Gly α59. The distances between non-hydrogen atoms (depicted by dotted lines) are given in Å.

Reproduced from Natarajan et al. (2018). See Chapter 9, section 9.5.3.

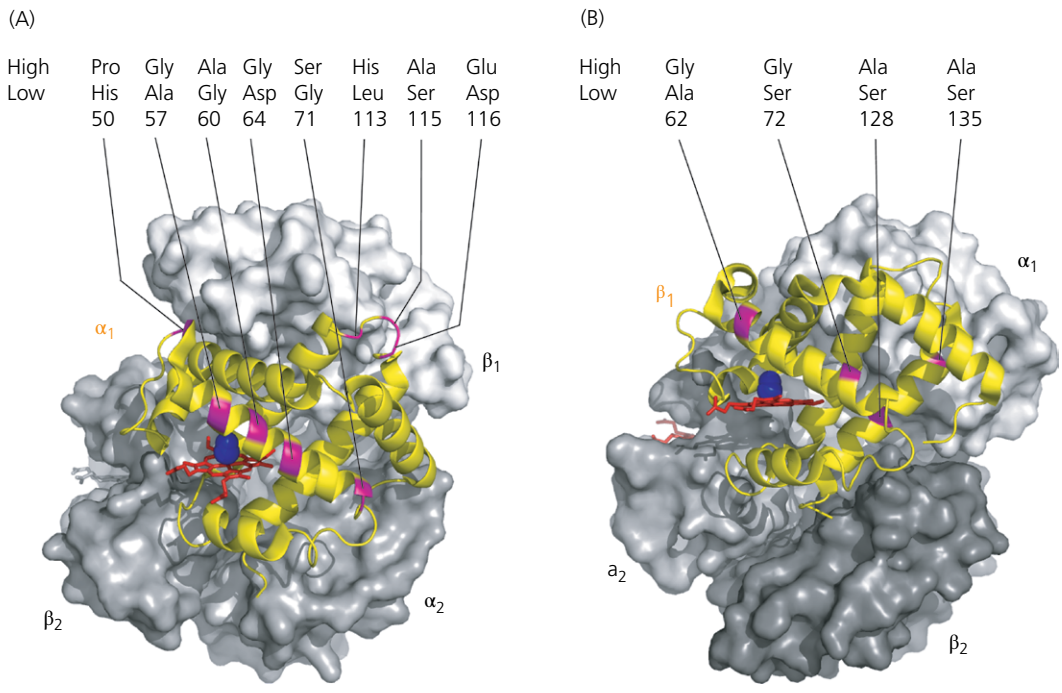

Plate 15. Amino acid polymorphisms in the Hbs of deer mice (*Peromyscus maniculatus*) contribute to genetic differences in Hb-O_2 affinity between high-altitude populations in the Rocky Mountains and low-altitude populations in prairie grassland. Eight sites in the α-chain (A) and four sites in the β-chain (B) exhibit striking altitudinal differences in allele frequency (Storz et al. 2010).

Reproduced from Storz et al. (2009). See Chapter 9, section 9.5.4.

CHAPTER 1

Principles of protein structure

1.1 Introduction

In the chapters to follow, we will explore the oxygenation properties of Hb and its physiological role in respiratory gas transport. Hb is a complex and exquisitely constructed molecular machine. The circulatory conveyance of chemically bound O_2 is its *raison d'être* and it also has physiologically important interactions with carbon dioxide (CO_2) and nitric oxide (NO). In each of the four subunits of the tetrameric Hb protein, a single O_2 molecule binds reversibly to an iron atom that is coordinated by four coplanar nitrogens at the center of a flat porphyrin ring (the heme group). Each heme is enclosed in a folded protein chain (the globin), to which it is bound via a coordinate covalent bond. How does the protein chain modulate O_2 binding by the iron atom? And what is the benefit of enclosing the iron atom in a protein cage in the first place? To provide a foundation for addressing these questions and others, we will first briefly review relevant principles of protein structure.

1.2 The hierarchy of protein structure

Proteins are polymers of amino acids that are linked in a specific linear sequence by peptide bonds. The 20 standard amino acids have different side chains with different physicochemical properties (Fig. 1.1), and these properties dictate how the polymer folds into a three-dimensional structure—the native conformation. The way in which the linear chain of amino acids spontaneously folds into a functionally intact, three-dimensional molecule, and the rules that govern this origami-like process, are foundational problems in the field of structural biology.

The primary structure of a protein refers to the linear sequence of amino acids comprising a single polypeptide chain. This is the only level in the hierarchy of protein structure that can be directly predicted from gene sequence. The genetic code indicates which particular amino acid is specified by each triplet of mRNA nucleotide bases (Fig. 1.2), and therefore serves as a Rosetta Stone for translating DNA sequence into amino acid sequence. The secondary structure of a protein refers to the spatial arrangement of residues that are close together in the linear sequence, and which give rise to regular, repeating patterns of hydrogen-bonded main chain conformations such as α-helices and β-sheets. The tertiary structure of a protein refers to the spatial assembly of helices and sheets and the interactions between them. It describes the three-dimensional folding pattern of the polypeptide chain. Finally, quaternary structure refers to the spatial arrangement of individual subunit polypeptides, and is therefore only applicable to multimeric proteins like Hb.

1.3 The peptide bond

"A protein molecule has the advantage of constructional simplicity that comes from being built from backbone parts of standardized dimensions."

—**Dickerson and Geis (1983)**

Each amino acid consists of a central carbon atom (called the α-carbon or "C_α") with an attached amino group (NH_3^+), a carboxylic acid group (COO^-), a

Hemoglobin: Insights into Protein Structure, Function, and Evolution. Jay F. Storz, Oxford University Press (2019).
 DOI: 10.1093/oso/9780198810681.001.0001

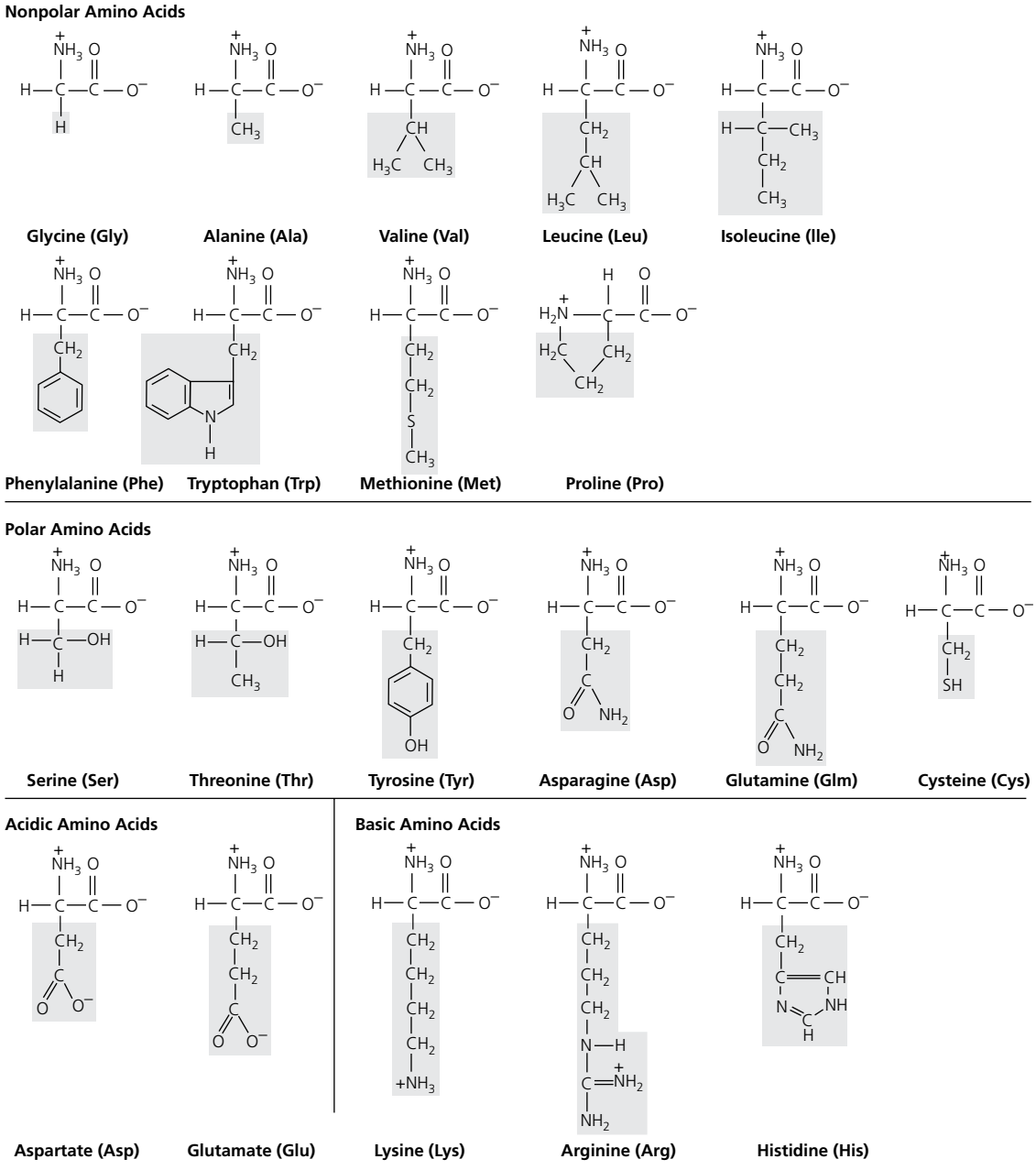

Fig. 1.1. The twenty standard amino acids in proteins. The pictured ionization states are those that predominate at pH 7. Side chains of each amino acid are indicated by shading.

Reproduced from McKee and McKee (2012) with permission from Oxford University Press.

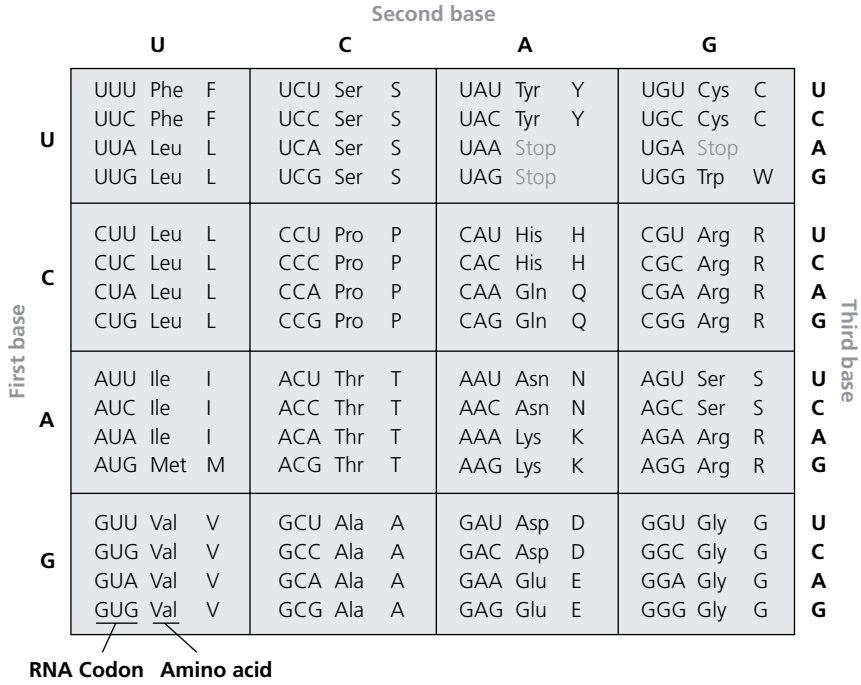

Fig. 1.2. The standard genetic code. The coding sequence of a gene specifies the order in which amino acids are linked together in the encoded protein, with each unique triplet of nucleotide bases (codon) specifying a particular amino acid or a punctuation mark (e.g., a stop codon that signals the end of the protein chain).

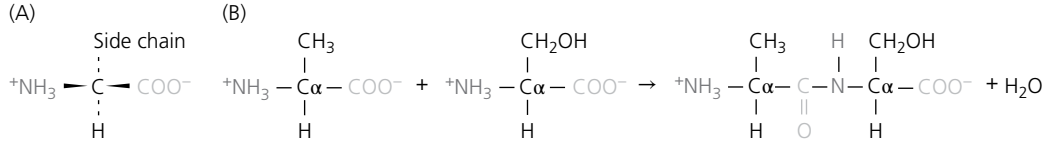

Fig. 1.3. Amino acids and the formation of peptide bonds. (A) General structure of an amino acid. (B) A peptide bond is formed when the α-carboxyl group of one amino acid reacts with the amino group of another, resulting in the elimination of a water molecule. This example shows the formation of a dipeptide involving alanine (side chain = CH_3) and serine (side chain = CH_2OH). In the cell, the synthesis of peptide bonds is an enzymatically controlled process that occurs on the ribosome and is directed by the mRNA template.

Modified from Lesk (2010) with permission from Oxford University Press.

hydrogen atom, and a characteristic side chain (Fig. 1.3A). Amino acids are covalently linked together in a protein chain via peptide bonds that are formed by the reaction between the carbon atom of the carboxylic acid group of amino acid "n" and the amino group of amino acid "n + 1," resulting in the loss of a water molecule (Fig. 1.3B). Individual amino acids are linked together end-to-end to form the main chain or backbone of a protein polypeptide with 20 different possible types of side chain protruding from the α-carbon of each residue (Fig. 1.4). The amino group of the first residue in the chain and the carboxylic acid group of the last residue remain intact, so the polypeptide chain is described as extending from the amino (N) terminus to the carboxy (C) terminus. Individual residues in the chain are numbered accordingly, from the N- to the C-terminus. This is also the order in which polypeptides are synthesized at the ribosome, as each new amino acid is added to the free carboxy terminus of the growing chain.

The stability and polarity of peptide bonds is attributable to resonance, the delocalization of electrons over adjoining chemical bonds (Martin, 2001, Voet and Voet, 2011). Because of resonance,

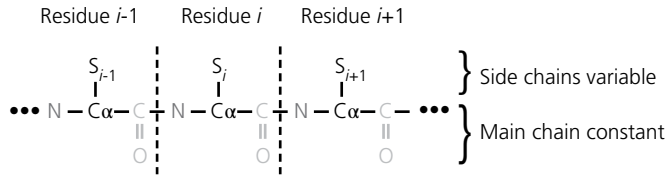

Fig. 1.4. Proteins are polymers of amino acids containing a constant main chain of repeating units with variable side chains. It is the sequence of variable amino acid side chains that gives each protein its distinctive character.

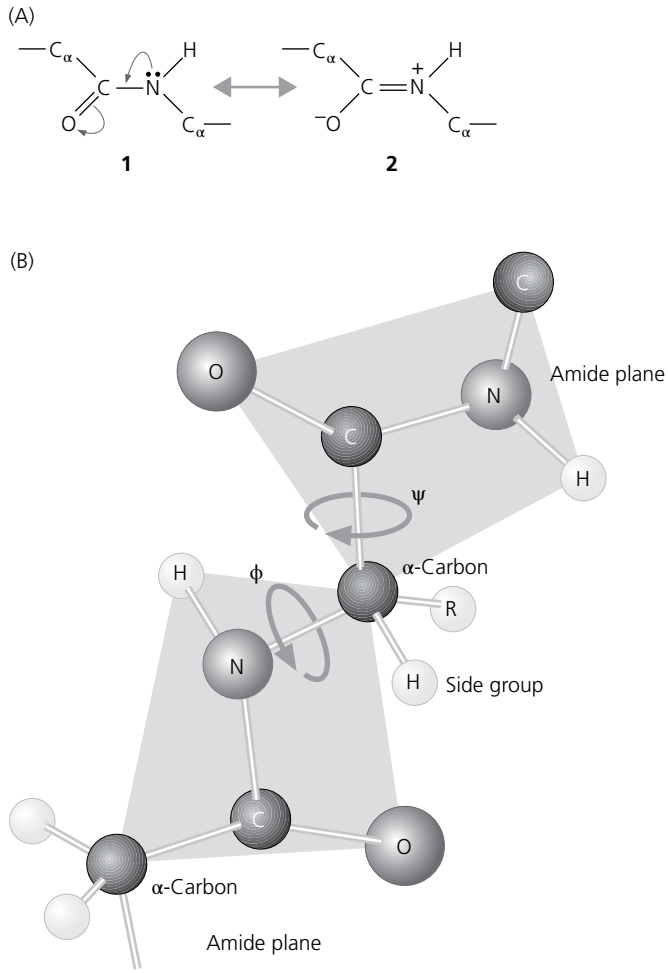

Fig. 1.5. The peptide bond. (A) Resonance forms of the peptide bond. (B) The partial double-bond character of the C-N bond means that all of the atoms connected by the shaded quadrilateral lie in the same two-dimensional plane, called the amide plane. Due to the rigidity of the peptide bond, the polypeptide chain can only rotate around the Cα—C bond (with a rotation angle of ψ) and the Cα—N bond (with a rotation angle of ϕ). This limits conformational degrees of freedom of the polypeptide chain, and therefore dictates allowable folding patterns in three dimensions.

Reproduced from McKee and McKee (2012) with permission from Oxford University Press.

the double-bond character of the C=O bond is shared with the adjoining C-N bond (Fig 1.5A), thereby preventing free rotation about that bond. Consequently, the three non-hydrogen atoms that make up each individual peptide bond (the carbonyl oxygen O, the carbonyl carbon C, and the amide nitrogen N) lie in the same two-dimensional "amide plane." Consecutive amide planes can only rotate about the single N-Cα and Cα-C bonds, so the α-carbons represent swivel points in the polypeptide chain. The angle of rotation around the N-Cα bond is called the Φ (phi) torsion angle and that around the Cα-C bond is called the Ψ (psi) torsion angle (Fig. 1.5B**)**. The allowable angles of rotation around the N-Cα and Cα-C bonds are constrained by steric hindrance between main chain atoms and side chain atoms. Thus, rotatable (but sterically constrained) covalent N-Cα and Cα-C bonds alternate with comparatively rigid peptide bonds along the main chain, and this imposes limits on the number of possible folded conformations a polypeptide chain can adopt (Ramachandran et al., 1963, Richardson and Richardson, 1990).

1.4 Folded proteins are mainly stabilized by weak, non-covalent interactions

The main chain polypeptide is linked together by covalent bonds, but the three-dimensional structure of native state proteins is mainly stabilized by a multitude of non-covalent, weakly polar interactions (Burley and Petsko, 1988, Jaenicke, 2000). These weakly polar interactions depend on the electrostatic attraction between opposite charges. The strength of association between positively and negatively charged atoms or groups of atoms depends on the distance between them and whether the interaction involves full or partial charges.

Van der Waals interactions involve a weak attractive force between atoms caused by fluctuations in electron density around their nuclei (Fig. 1.6A). The interaction is strongest between groups that are the most polarizable, such as the methyl groups and methylene groups of hydrophobic amino acid side chains. Van der Waals interactions are exclusively short range, generally involving atoms less than 5 Å apart.

Hydrogen bonds are formed between atoms of nitrogen (N) and oxygen (O) via an intermediate hydrogen atom (H): For example, N-H…N, NH…O, OH…O, or OH…N, where in each case the hydrogen is covalently bonded to the atom on the left (the donor) and more weakly bonded to the negatively polarized, non-bonded atom on the right (the acceptor) (Fig. 1.6B). The most common hydrogen bonds in proteins involve the N-H and C=O groups of the polypeptide main chain. In this hydrogen bond, N-H…O=C, the typical H…O distance is 1.9–2.0 Å, whereas the covalent N-H distance is ~1 Å. The hydrogen atom that is covalently bound to the more electronegative donor atom has a partial

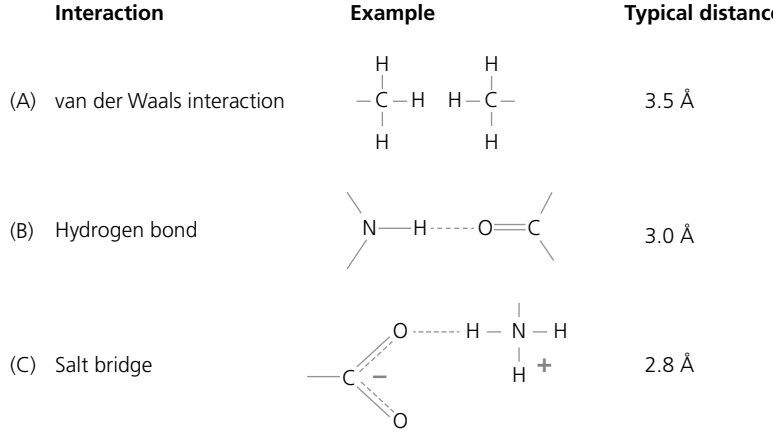

Fig. 1.6. Summary of non-covalent interactions that stabilize polypeptides. Interatomic distances for the different interactions are highly context-dependent, so the values shown here should be regarded as approximate averages.