Structure, Vol. 10, 725–735, May, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00763-3

The 109 Residue Nerve Tissue Minihemoglobin from *Cerebratulus lacteus* Highlights Striking Structural Plasticity of the α-Helical Globin Fold

Alessandra Pesce,[1] Marco Nardini,[1] Sylvia Dewilde,[2] Eva Geuens,[2] Kiyoshi Yamauchi,[3] Paolo Ascenzi,[1,4] Austen F. Riggs,[5] Luc Moens,[2] and Martino Bolognesi[1,6]

[1]Department of Physics
INFM
Advanced Biotechnology Centre
University of Genova
Largo Rosanna Benzi 10
I-16146 Genova
Italy

[2]Department of Biochemistry
University of Antwerp
Universiteitsplein 1
B-2610 Antwerp
Belgium

[3]Department of Biology and Geoscience
Faculty of Science
Shizuoka University
836 Oya
Shizuoka 422-8529
Japan

[4]Department of Biology
University "Roma Tre"
Viale Guglielmo Marconi 446
I-00146 Roma
Italy

[5]Section of Neurobiology and
Institute of Cell and Molecular Biology
School of Biological Sciences
University of Texas
CO920
Austin, Texas 78712

Summary

A very short hemoglobin (CerHb; 109 amino acids) binds O_2 cooperatively in the nerve tissue of the nemertean worm *Cerebratulus lacteus* to sustain neural activity during anoxia. Sequence analysis suggests that CerHb tertiary structure may be unique among the known globin fold evolutionary variants. The X-ray structure of oxygenated CerHb (R factor 15.3%, at 1.5 Å resolution) displays deletion of the globin N-terminal A helix, an extended GH region, a very short H helix, and heme solvent shielding based on specific aromatic residues. The heme-bound O_2 is stabilized by hydrogen bonds to the distal TyrB10-GlnE7 pair. Ligand access to heme may take place through a wide protein matrix tunnel connecting the distal site to a surface cleft located between the E and H helices.

Introduction

Hemoglobins (Hbs) occur in all kingdoms of living organisms and display a variety of functions. In addition to O_2 transport and storage [1–3], several novel Hb functions have recently emerged, including control of NO levels in microorganisms and dehaloperoxidase activity [4–7]. Despite the large variability in their primary and quaternary structures, Hbs display a well-conserved tertiary structure (the "globin fold"), consisting of 140–160 amino acids typically organized in a three-on-three α-helical sandwich [8–10].

Hb-like molecules are found intracellularly or dissolved in the circulating body fluid (hemolymph). Three categories can be distinguished within the intracellular or cytoplasmic globins: Hbs located in circulating erythrocyte-like cell types, the tissue Hbs, and the Hbs of the unicellular organisms [7]. Examples of the first category have been largely represented by many functional and structural studies on vertebrate blood Hbs over the years [11]. Monomeric myoglobin (Mb) is the best known tissue globin; it occurs in muscles, where it acts as an O_2 buffer facilitating oxygen diffusion to the mitochondria [2, 3], also involved in NO detoxification and biochemistry [12]. Unicellular organisms may display different globin subfamilies, including full-length Hbs, truncated Hbs (trHbs), and the chimeric flavohemoglobins [4, 10, 13, 14].

Nerve tissue Hbs are found in vertebrates and invertebrates. Neuroglobin and cytoglobin are recently discovered new members of the vertebrate globin family. Neuroglobin is a monomeric, ~150 amino acid-long heme protein, predominantly expressed in vertebrate brain at micromolar concentration, displaying less than 25% sequence identity to conventional vertebrate Hbs or Mbs (see Figure 1) [15, 16]. Its O_2 affinity is comparable to that of Mb (P_{50} = 2 torr), with a heme hexacoordinated structure of the His-Fe-His type in both deoxyferrous and -ferric forms. The function of neuroglobin is still debated [5, 15, 17–19]. Cytoglobin is thought to be expressed at very low concentrations in all vertebrate tissues, including brain [16]. It displays sequence identities versus Hbs and Mbs of about 26%–30%, being characterized by N- and C-terminal extensions of 15–20 amino acids. Thus, despite occurring also in the brain, cytoglobin cannot be classified as nerve specific. However, both human neuroglobin and cytoglobin genes display unique intron/exon patterns; phylogenetic analysis suggests a very ancient origin [15–17].

Nerve tissue globins have sporadically been observed in mollusc, annelid, arthropod, and nemertean species as well as in nematodes [2, 20–22]. In contrast to their vertebrate counterparts, the invertebrate nerve tissue globins reach a millimolar local concentration. This high concentration is likely sufficient to facilitate O_2 diffusion or storage, thus supporting cell function during temporary hypoxia periods encountered by the animal [2, 21, 23, 24].

From a structural viewpoint, sequence alignment of

[6]Correspondence: bolognes@fisica.unige.it

Key words: globin fold plasticity; nerve tissue minihemoglobin; oxygen binding; oxygenated hemoglobin; protein cavities; X-ray crystallography

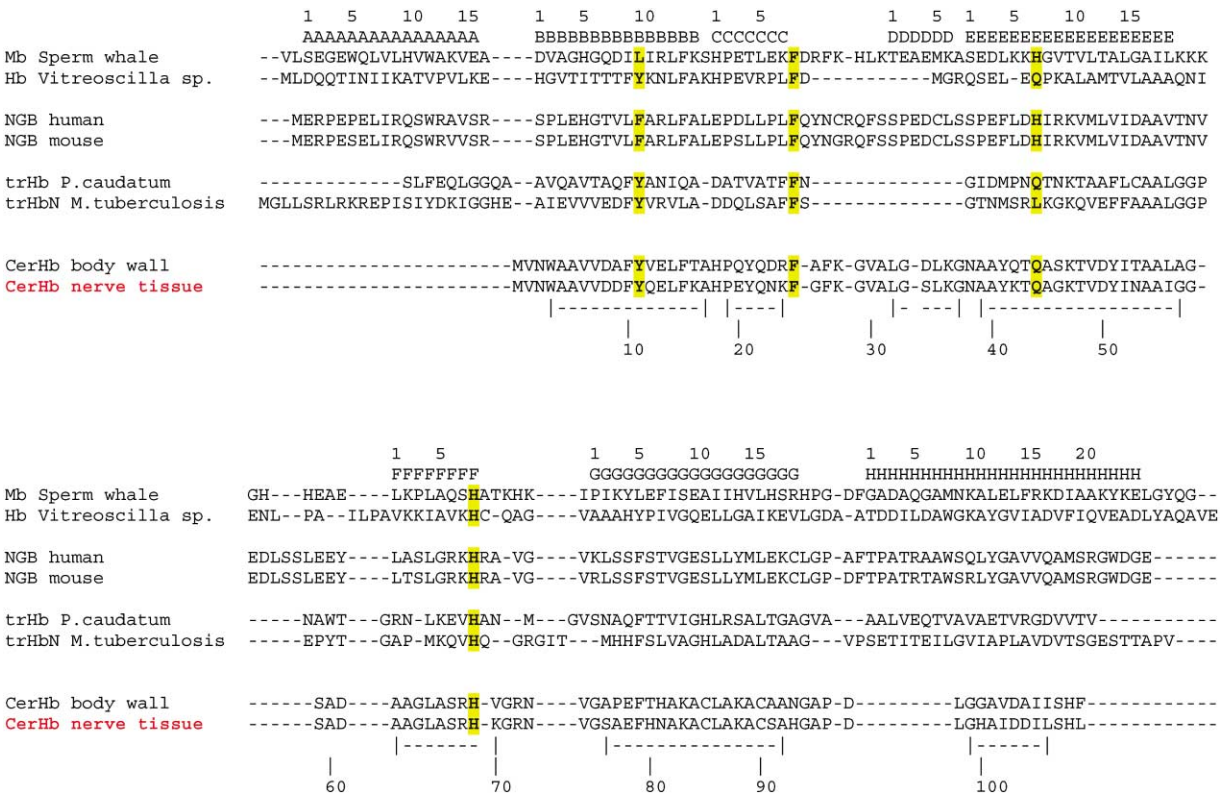


Figure 1. Structure-Based Amino Acid Sequence Alignments

The figure displays multiple amino acid sequence alignments of CerHb relative to *C. lacteus* body wall Hb, *P. caudatum* trHb, and *M. tuberculosis* trHbN (as representatives of the trHb homology subfamily), sperm whale Mb, *Vitreoscilla* sp. Hb, and human and mouse neuroglobins (NGB). With the exception of human and mouse neuroglobins, the sequence alignments have been based on structural superpositions of each protein's three-dimensional structure onto the CerHb Cα backbone. The extension of α helices in sperm whale Mb has been indicated on top of the alignment, with reference to topological sites. Amino acid numbering and the span of α helices in CerHb are shown on the bottom line. The key topological positions B10, CD1, E7, and F8 are highlighted in yellow.

all known nerve tissue globins reveals the presence of key globin fold determinants [15, 25, 26]. In general, most globin α-helical segments can be recognized, together with the vertebrate globin invariant residues PheCD1, HisF8, and patterns of buried hydrophobic residues (Figure 1) [25–27]. (Amino acid residues have been identified by their three-letter code, their sequence number, and their topological position within the eight helices—A through H—of the globin fold [1]). However, in specific cases such as the nerve tissue and body wall Hbs of *C. lacteus*, which are the smallest stable Hbs known (both composed of 109 amino acids), extensive perturbation of the common globin fold pattern is evident together with poor conservation even of the structural determinants typical of the truncated Hb subfamily [13, 14, 28].

Here we describe the three-dimensional structure of oxygenated *C. lacteus* nerve tissue Hb (CerHb) at 1.5 Å resolution, as a first structural insight into the newly discovered neuroglobin subfamilies. CerHb tertiary structure displays striking editing of the vertebrate globin fold, with deletion of the N-terminal A helix and shortening of the C-terminal H helix to two turns. A unique protein matrix tunnel, connecting the heme distal cavity to the molecular surface, may play a central role in supporting O_2 diffusion to and from the heme.

Results and Discussion

Deviations from the Classical Globin Fold in the 109 Residue CerHb

The three-dimensional structure of the oxygenated derivative of *C. lacteus* nerve tissue Hb (CerHb) was solved by means of MAD methods, based on the anomalous signal of the heme iron atom. Diffraction data were collected at three wavelengths at the ID29 ESRF beamline (Grenoble, France; Table 1), using one crystal of the orthorhombic form previously described [29]. Refinement of the crystal structure converged at a general R factor value of 15.3% (R_{free} 18.7%) for data in the 35.0–1.5 Å resolution range, with ideal stereochemical parameters [30, 31]. The final model contains 857 protein atoms, 106 ordered solvent atoms, one dioxygen molecule, and one sulfate and one acetate anions (Table 1).

Despite the low levels of residue conservation relative to full-length (non)vertebrate Hbs or Mbs (the residue identity between sperm whale Mb and CerHb aligned sequences is 13%), CerHb can be predicted to fall within the Hb superfamily (Figure 1) [1, 9, 10, 21, 27]. However, inspection of the crystal structure shows that CerHb is strikingly modified at two main regions, due to extensive residue deletions relative to full-length globins (Figures 2A and 2B). First, the N-terminal A helix of the classical

Table 1. Data Collection and Refinement Statistics for CerHb

(A) CerHb MAD Data Collection Statistics

	Absorption peak	Inflection point	Remote
Wavelength (Å)	1.739	1.740	0.915
Resolution (Å)	35–1.95	35–1.95	35–1.50
Mosaicity (°)	0.42	0.42	0.39
Completeness (%)	93.4 (90.2)*	93.5 (89.9)	97.1 (95.7)
R_{merge} (%)	4.8 (8.5)	4.9 (8.9)	5.4 (23.9)
Total reflections	29,384	29,367	67,606
Unique reflections	8,559	8,544	17,906
Redundancy	3.4	3.4	3.8
Average I/σ(I)	16.7 (10.2)	16.3 (9.4)	14.3 (3.2)

(B) Refinement Statistics and Model Quality

Resolution range (Å)	35–1.50
Protein nonhydrogen atoms	857
Water molecules	106
Dioxygen molecules	1
Sulfate ions	1
Acetate ions	1
R factor/R_{free}# (%)	15.3/18.7
Space group	$P2_12_12_1$
Unit cell (Å)	$a = 42.72$, $b = 43.17$, $c = 60.11$
Rmsd from ideal geometry:	
Bond lengths (Å)	0.006
Bond angles (°)	1.14
Ramachandran plot:##	
Most favored region	95.7%
Additional allowed region	4.3%
Averaged B factors (Å^2):	
Main chain	14
Side chain	16
Solvent	24
Heme	13

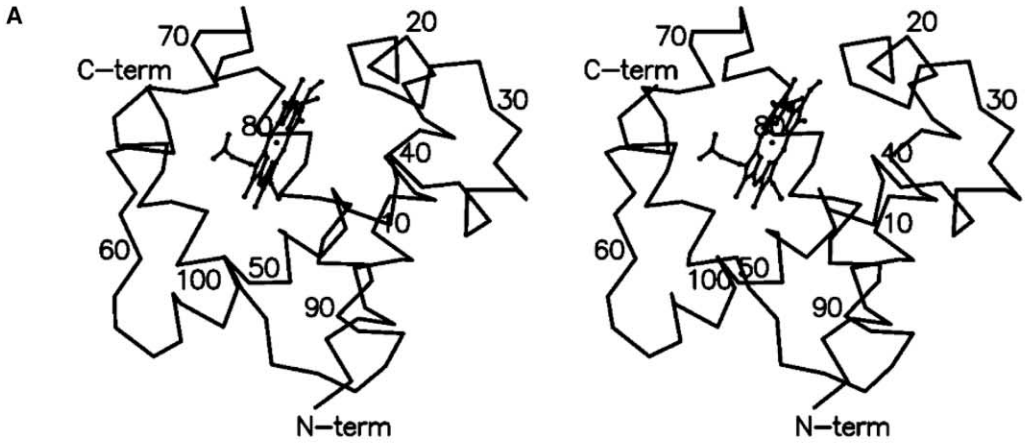
*Outer shell statistics are shown in parentheses. The outer shells are 1.95–2.00 Å for the absorption peak and inflection point, and 1.53–1.50 Å for the remote point.
#Calculated using 10% of the reflections.
##Data produced using the program PROCHECK [31].

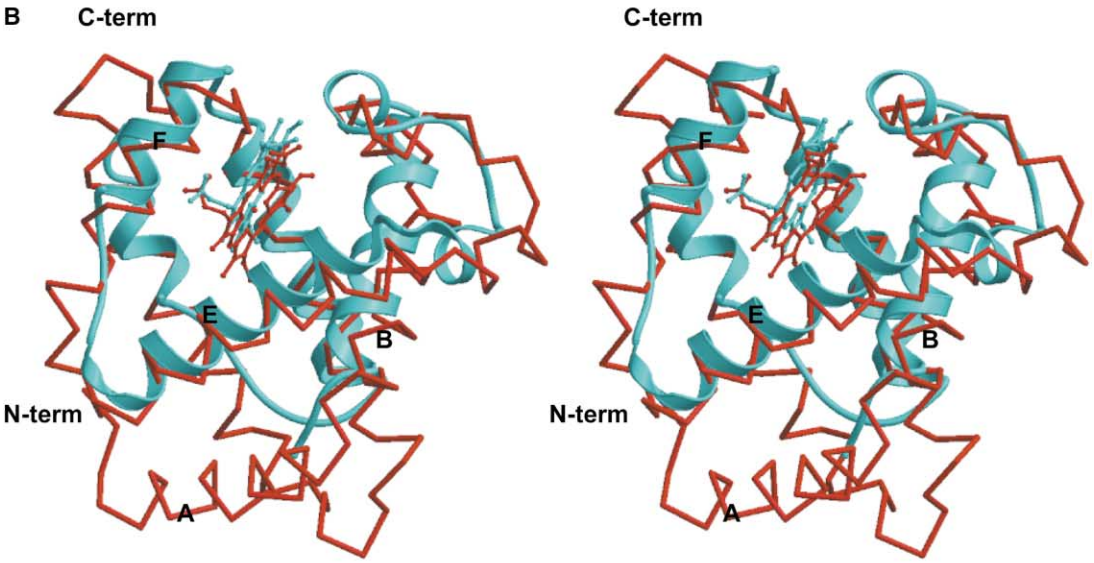
globin fold is entirely deleted. Second, the (C-terminal) H helix is trimmed to only 8 residues [Gly(99)H10–Leu(106)H17], being largely substituted by an extended GH interhelical region. Thus, from a globin topological viewpoint, CerHb is essentially composed of helices B through H of the classical globin fold (Figure 2).

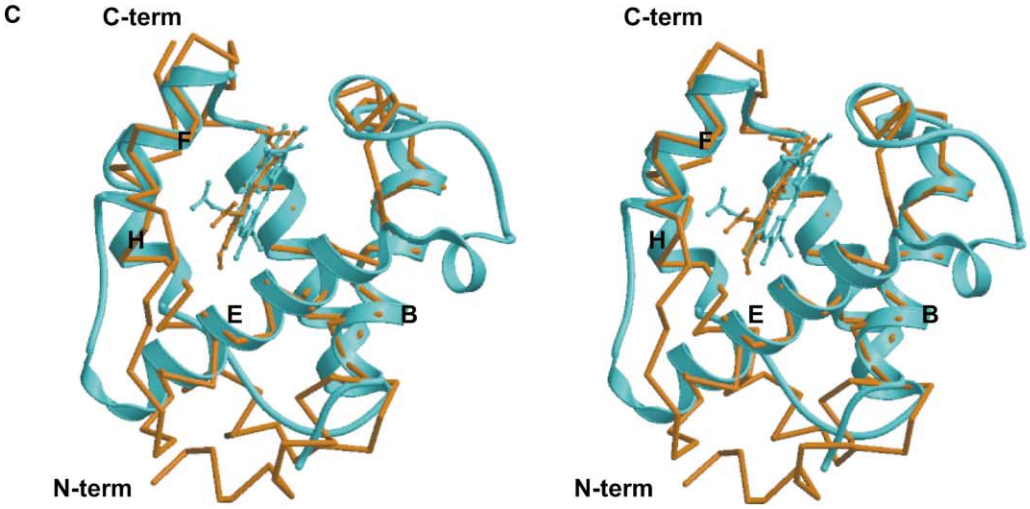
Within the Hb superfamily, CerHb [21] and truncated Hbs [13, 28] display the lowest known molecular weights (12–13 kDa). However, a structural overlay of CerHb (109 amino acids) and *P. caudatum* trHb (116 amino acids) protein backbones (Figure 2C) highlights the matching of only 67 Cα pairs (rmsd = 1.33 Å), mostly localized at the B, E, and G helices. The poorly matching regions in the two globins fall on the heme proximal side, particularly along the F helix (an extended polypeptide pre-F region in trHbs), in keeping with the low sequence identity (<20%) relating CerHb and the trHb subfamily (Figures 1 and 2C). Similarly, despite the substantial deletions of residues mentioned above, Cα backbones of CerHb and sperm whale Mb match at 68 Cα pairs, with an rmsd of 1.77 Å. Further, as a general rule emerging from different structural comparisons [32, 33] of CerHb within the Hb superfamily, a net preference for conservation of the full B-, C-, E-, and G-helical regions among all (non)vertebrate globins is evident. Thus, based on fold comparisons only, CerHb appears almost equally distant from all known globin tertiary structures, supporting the identification of a new Hb subfamily (the "mini-Hb" subfamily [21]).

Structural Features Characteristic of the CerHb Fold

As a result of the A helix deletion (18 residues deleted relative to sperm whale Mb; see Figures 1 and 2B), CerHb N-terminal residue Val1 falls at the AB hinge region within the classical globin fold topology. In the absence of the whole A helix, a globin-conserved and stabilizing intramolecular hydrophobic contact between residue A12 (mostly Trp [25]) and residues from the C-terminal end of the E helix is lost, partly substituted in CerHb by contacts between Trp(3)B2 and Val(1), Ile(52)E15, Ile(56)E19, and the GH region residues Ala(95) and Leu(98) (Figure 3). Such a localized hydrophobic cluster anchors the B helix to the CerHb core, and efficiently seals the heme distal pocket from solvent access through a B/E and B/G contact region route, the primary route proposed for ligand diffusion to the heme in trHbs [28]. The occurrence of Trp at the B2 topological position is extremely uncommon throughout all globin sequences [25, 26, 34], being present only in *Lucina pectinata* HbI, where the surrounding structural region is rich in aromatic residues [35]. In CerHb, the occurrence


A
70
20
C–term
30
80
40
40
10
60
50
100
90
N–term


B
C-term
C-term
F
E
B
N-term
A


C
C-term
C-term
F
H
E
B
N-term

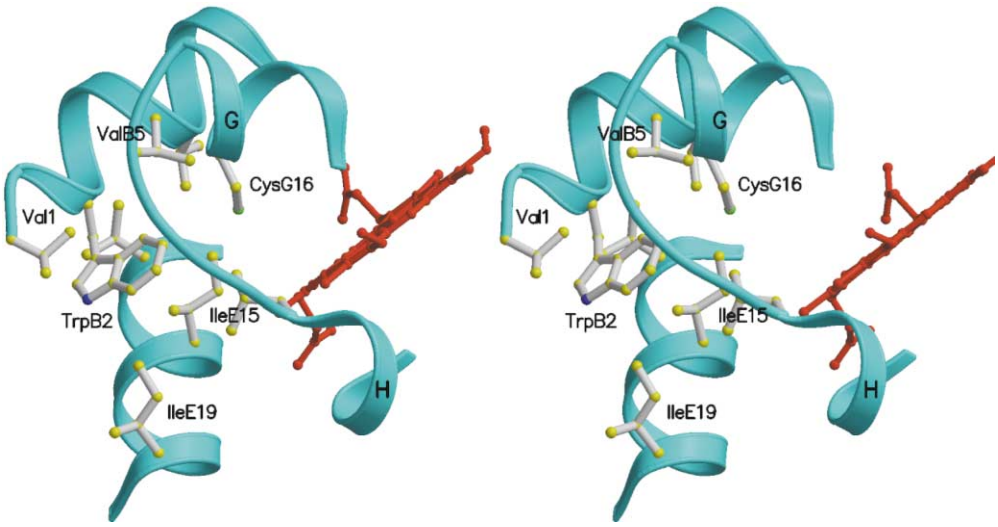


Figure 3. The Structural Environment of Trp(3)B2

The CerHb elongated GH segment (shown in foreground) appears to be related to the presence of the bulky residue Trp(3)B2, which anchors the N-terminal part of the molecule to the protein core through several hydrophobic contacts. Notice also the location of residue Cys(90)G16, relative to the heme. Drawn with MOLSCRIPT [66] and Raster3D [67].

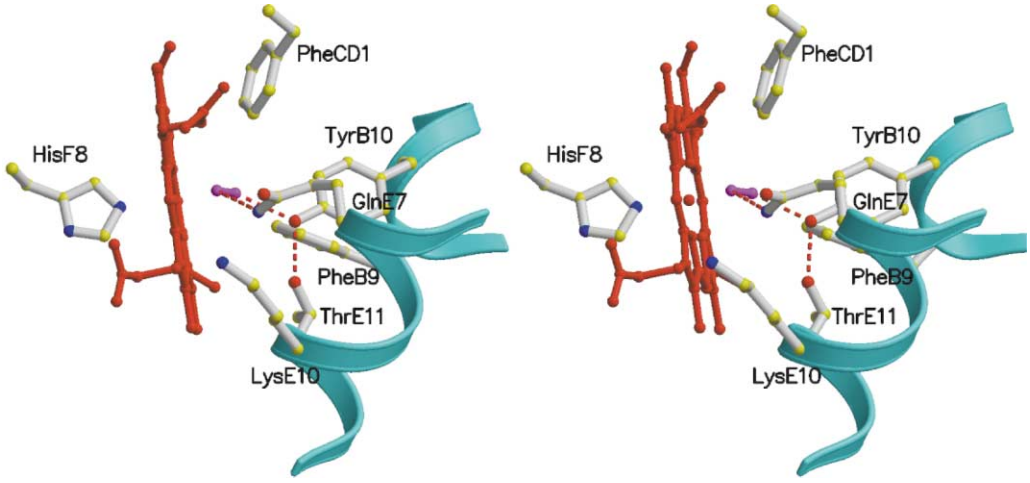


Figure 4. O_2 Stabilization at the Heme Distal Site of CerHb

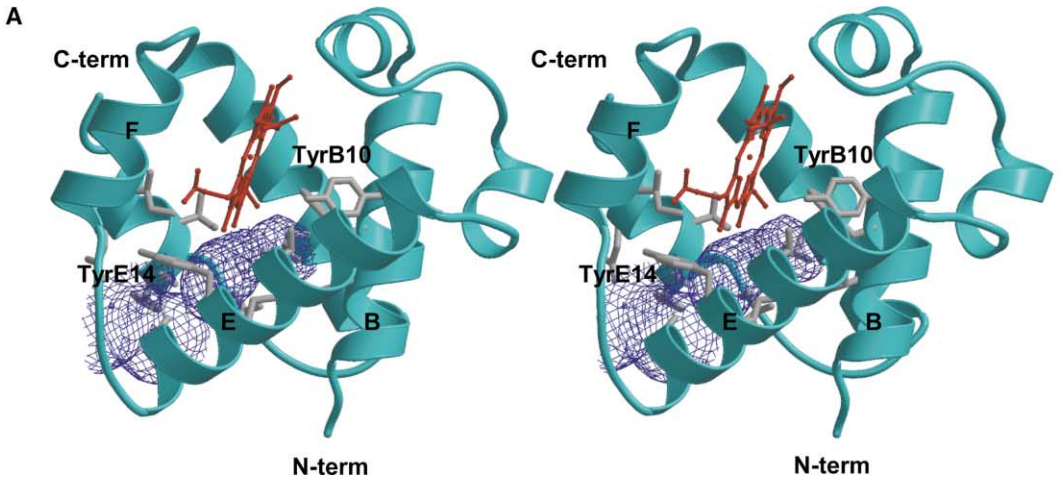
The CerHb distal site region is displayed in a stereo view, showing residues Phe(10)B9, Tyr(11)B10, Phe(25)CD1, Gln(44)E7, Lys (47)E10, and Thr(48)E11, the dioxygen molecule (magenta), the heme group, and part of the polypeptide chain path. Hydrogen bonds are drawn as dashed lines. The proximal HisF8 residue is also shown. Drawn with MOLSCRIPT [66] and Raster3D [67].

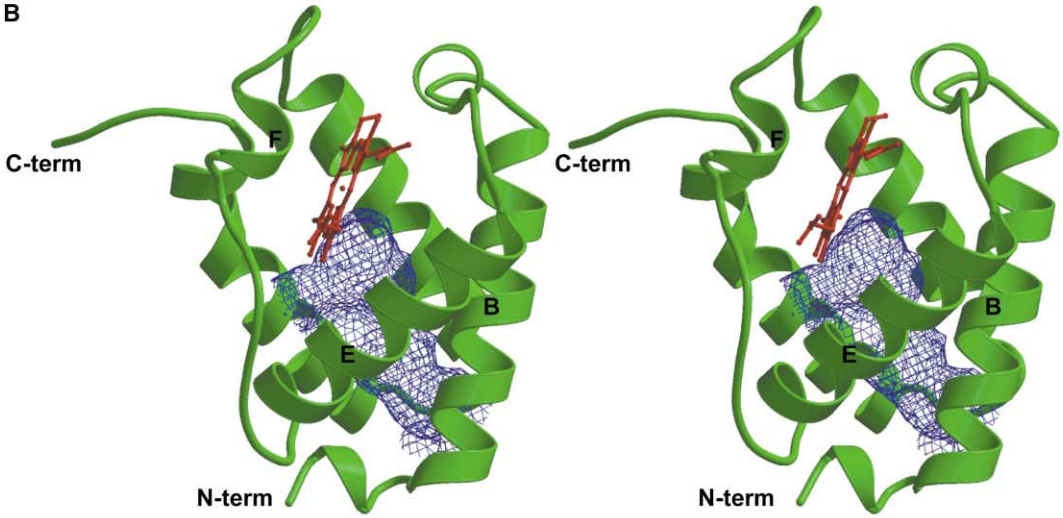
of a bulky Trp residue at site B2 close to the GH region within the tertiary structure likely restricts the presence of an α-helical segment at the N-terminal half of the H helix which is substituted by the extended GH polypeptide segment His(93)–Leu(98) (Figures 2 and 3). To prevent formation of an α-helical segment at this topological region, the GH segment has adopted the helix-breaking Gly(94)-Pro(95) sequence motif.

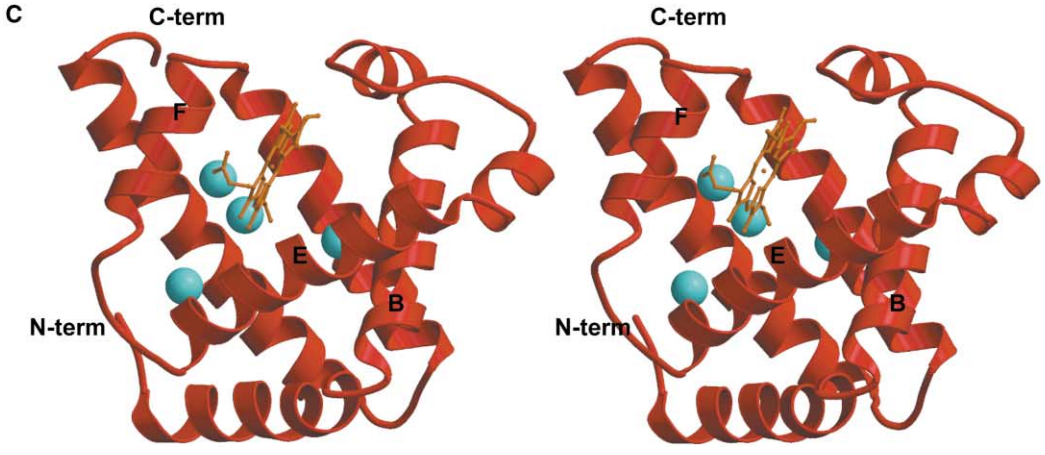
Inspection of the CerHb core regions highlights structural details which are reminiscent of (non)vertebrate globins. The B helix supports the Phe(10)B9-Tyr(11)B10 residue pair, setting the orientation of Tyr(11)B10 toward the distal heme site (see below). A similar residue pair is present in elephant Mb [36], and is strongly conserved in nonvertebrate globins [7, 10, 26, 34]. As observed in several vertebrate globins [25], the His(18)C1-Pro(19)C2

Figure 2. CerHb Tertiary Structure

Stereo views of CerHb Cα trace ([A]; every tenth residue and the N- and C-terminal residues are labeled). CerHb Cα backbone (cyan ribbon), in overlays with sperm whale Mb (red trace, [B]; the main α helices are labeled in sperm whale Mb) and *P. caudatum* trHb (orange trace, [C]; α helices are labeled in CerHb). Drawn with MOLSCRIPT [66] and Raster3D [67].


A
C-term
F
TyrB10
TyrE14
E
B
N-term
C-term
F
TyrB10
TyrE14
E
B
N-term


B
C-term
F
B
E
N-term
C-term
F
B
E
N-term


C
C-term
F
E
B
N-term
C-term
F
E
B
N-term

pair of CerHb starts the C helix, which supports residue Phe(25)CD1 next to the heme CHD methinic bridge. Structural similarity between CerHb and sperm whale Mb, in this region, is substantial (Figure 2B).

The CerHb CD region displays contacts between Phe(25)CD1 and Phe(27)CD3 side chains. Similar stacking interactions are observed in (non)vertebrate globins, although the sequence motif is often 1 residue longer, that is, PheCD1-X-X-PheCD4. Notable residues along the CerHb E helix are Gln(44)E7 and Thr(48)E11, which face the heme distal ligand binding site, and Tyr(51)E14, which provides a floor to the heme pocket, compensating for the absence of the A helix. A similar function for a strongly conserved Phe residue at site E14 has been noted in trHbs, where virtual absence of the proximal F helix would leave a solvent-accessible heme cavity [13, 28].

Gly(57)E20 is the last residue in the E helix, immediately followed by the Gly(58)–Asp(61) EF hinge. The heme proximal F helix covers residues Ala(62)F1 through Lys(70)F9, encoding a heme proximal side overall structure more closely related to (non)vertebrate globins than to trHbs, which also display a conserved Gly-Gly motif at the EF hinge region [13, 28]. CerHb F helix regularly supports residue Leu(65)F4, contacting the heme CHB methinic bridge and the proximal His(69)F8 residue.

The elongated G helix (15 amino acids) of CerHb hosts residue Phe(79)G5, next to the heme B-pyrrole ring and to the imidazole ring of His(69)F8. An intramolecular contact between Phe(79)G5 and His(69)F8 sets the azimuthal orientation of the proximal imidazole relative to the heme pyrrole N atoms. The G helix hosts 2 Cys residues, at sites G11 and G16. Both residues are too far from the distal site cavity to suggest any involvement in heme ligand reactivity (i.e., O_2/NO chemistry), as proposed for *Ascaris suum* Hb [37].

Heme Stabilization and Ligand Binding

Stabilization of the heme group within the CerHb polypeptide chain occurs through 35 van der Waals contacts [$\leq$4.0 Å; residues Phe(10)B9, Tyr(21)C4, Lys(24)C7, Tyr(51)E14, Leu(65)F4, Arg(68)F7, Arg(72), Val(74), Phe(79)G5, Ala(82)G8, and Ile(102)H13], and through the His(69)F8 NE2-Fe coordination bond. Moreover, electrostatic interactions to the heme propionates provide a significant stabilizing contribution. Residue Arg(72) is hydrogen bonded to the heme D-propionate, whereas Lys(47)E10 is electrostatically linked to the heme A-propionate (at 4.9 Å). The side chain of Arg(68)F7 may also contribute to electrostatic stabilization of the heme, but in the crystal structure its orientation is affected by the presence of a nearby bound sulfate anion.

Binding of O_2 to the heme Fe atom occurs in a tightly occupied distal site cavity (Figure 4). The O_2 molecule is bound to the heme Fe atom via a 1.94 Å coordination bond. The dioxygen molecule adopts a bent orientation, forming a 103° Fe-O1-O2 angle. Relative to the porphyrin ring, the O_2 ligand is oriented in the direction of the methinic CHD atom, thus pointing to the rear end of the heme crevice. Two main hydrogen bonds stabilize the heme Fe-bound dioxygen. On the one hand, the Fe distal O2 atom is linked to the Tyr(11)B10 OH group (2.58 Å). On the other, the Fe proximal O1 atom is hydrogen bonded to the Gln(44)E7 NE2 atom (2.61 Å). Additional polar interactions contribute to the overall structural organization of the heme distal site. Specifically, the Tyr(11)B10 OH group is connected to the Thr(48)E11 OG1 atom by a strong hydrogen bond (2.59 Å), while weaker interactions connect the Tyr(11)B10 OH group to the Gln(44)E7 NE2 atom (3.24 Å), and the Gln(44)E7 NE2 atom to the Thr(48)E11 OG1 atom (3.53 Å). Moreover, it can be noticed that a close contact (3.5 Å) occurring between the heme Fe-bound dioxygen and the rim of the Phe(25)CD1 aromatic ring is indicative of an aromatic-electrostatic contact, which further contributes to stabilization of the ligand (Figure 4) [38]. As a result of the interlacing of side chains and hydrogen bonds, the heme Fe-bound dioxygen is fully buried in the distal site, being totally inaccessible to solvent.

The side chain clustering within the heme distal site does not leave room for solvent molecules, in keeping with the absence of electron density peaks besides dioxygen in this protein region. Analysis of the atomic B factors for the O_2 molecule (O1 B = 23 Å^2, O2 B = 28 Å^2) is in keeping with a bound dioxygen displaying only slightly higher mobility at the distal O2 atom, hydrogen bonded by residue Tyr(11)B10. Such an observation, and the orientation adopted by the O_2 diatomic molecule, may reflect specific structural properties of the heme distal cavity. Spectrophotometric analysis of dissolved crystals (aged, but not exposed to X-rays) did not show the presence of deoxygenated or oxidized CerHb species, in keeping with the reported stability of oxygenated CerHb, in vivo and in vitro [21]. However, the occurrence of radiation-dependent redox processes during X-ray data collection cannot be ruled out [39].

The proximal His(69)F8 residue is coordinated to the heme Fe atom through a 2.06 Å bond. The proximal imidazole azimuthal orientation is set by a hydrogen bond between the side chain ND1 atom and Leu(65)F5 O atom (2.85 Å), and by contacts to Phe(79)G5. The F5-F8 main chain/side chain hydrogen bond is a key interaction strongly conserved in Hbs and Mbs from all evolutionary phyla [10]. Analysis of the heme Fe atom coordination parameters indicates that the Fe atom lies within the pyrrole N-atom's plane, with coordination distances to the four pyrrole N-atoms averaging 2.03 Å.

A Proposed Novel O_2 Diffusion Pathway in CerHb

Lack of the N-terminal A helix together with other CerHb-specific structural features such as residue substitutions

Figure 5. An Elongated Protein Matrix Cavity in CerHb

Outline of the CerHb core cavity, stretching from protein surface to the heme distal site. The cavity surface, defined by a 1.4 Å radius probe, is calculated in the absence of the O_2 molecule with the program SURFNET [68] and is displayed in blue. Residues lining the cavity walls are portrayed in light gray and the heme group in red (A). For comparison, *C. eugametos* trHb (B) and sperm whale Mb (C) are shown in the same orientation and scale, together with the protein tunnel surface and Xe binding sites (cyan spheres), respectively. Drawn with BOBSCRIPT [69] and Raster3D [67].

and the span of individual helices gives rise to an evident surface cleft, located approximately between the end of the E helix and the beginning of the H helix. This cleft is about 8 Å wide, bordered by surface residues His(100)H11, Asp(104)H15, and by main chain carbonyl groups of the 55–62 region, hosting several ordered water molecules. The inner part of the cleft is defined by residues Ala(55)E18, Leu(98)H9, and Ala(101)H12, which in turn are part of the narrow access to an elongated core tunnel (about 10 Å long, >100 Å^3 volume, as defined by a 1.4 Å probe) directed toward the heme distal site, where it is terminated by Val(7)B6, Phe(10)B9, and Thr(48)E11 (Figure 5A). The tunnel is mainly lined by 9 hydrophobic residues [Val(7)B6, Phe(10)B9, Ile(52)E15, Ala(55)E18, Leu(86)G12, Leu(98), Ala(101)H12, Ile(102)H13, Ile(105)H16], by Thr(48)E11 and Tyr(51)E14, and by the heme B-methyl and B-vinyl groups. The inner part of the tunnel has roughly an hour glass shape, divided between the proximal and distal sides of the heme, with a diameter varying between 5.5 and 6.9 Å. The heme distal side end of the tunnel falls at about 4 Å from the Tyr(11)B10 OH group and from the heme Fe-bound O_2 molecule. No other structural feature of the distal site limits access to the ligand binding site. The overall size and shape of the tunnel may vary dynamically, following conformational readjustments of the Tyr(51)E14 side chain, which, despite being a core residue, is relatively free of intramolecular contacts and can swing within the tunnel space by Cα-Cβ and Cβ-Cγ rotations, possibly reshaping it into two smaller regions.

CerHb core tunnel is unique in its size and location among all known globin structures. Considering the free energy cost of maintaining an empty cavity within the protein matrix relative to protein stability [40], it can be suggested that the CerHb core cavity plays a functional role in providing an O_2 diffusion pathway to and from the heme, as an alternative to the distal site gating mechanism based on the E7 residue found in classical Hbs and Mbs [41]. In fact, O_2 release and uptake to and from the solvent space through the CerHb distal E7 region would require a substantial conformational transition of the Gln(44)E7 side chain, likely associated with readjustments of the side chains of Lys(47)E10, the heme propionate(s), and cleavage of several intramolecular hydrogen bonds.

Related to the issue of ligand diffusion pathways to the heme, recent structural investigations on trHbs from *Mycobacterium tuberculosis* and *Chlamydomonas eugametos* have indicated that a 20 Å-long hydrophobic tunnel through the protein matrix may support O_2 diffusion and binding processes (Figure 5B) [13, 14, 28]. Moreover, local diffusion processes of diatomic ligands within the Mb protein matrix have shown the residence of O_2, CO, or NO in small internal cavities, previously identified by Xe atom binding [42]. Such processes have recently been related to Mb dynamics and functionality, opening new perspectives in the O_2-dependent NO detoxification process and reaction catalyzed by Mb in vivo [6, 43–47]. Inspection of the three-dimensional structure, however, shows that no internal cavities are present in CerHb besides the core hydrophobic tunnel. From the structural and topological viewpoints, CerHb core tunnel is unique and cannot be related to the protein matrix tunnel of trHbs, nor to the Xe cavities identified in Mb as relevant for ligand diffusion (Figure 5).

Functional and Molecular Properties of a Mini-Hb

In situ, CerHb is half-saturated with O_2 at 2.9 torr, the corresponding value of the apparent dissociation equilibrium constant for oxygen binding being 3.8×10^{-6} M (at pH 7.3 and 15°C). The Hill coefficient for CerHb in situ at half O_2 saturation, $n \cong 2.3$ (at pH 7.3 and 15°C), is in keeping with the self-association of this nerve tissue Hb at least into tetramers in the deoxygenated state at the 2–3 mM (heme) concentration estimated in the cells, dissociating into dimers and monomers upon oxygenation and dilution [21]. Therefore, CerHb undergoes O_2-linked association-dissociation equilibria [21] as often observed in (non)vertebrate globins [7, 10, 48]. A very similar behavior has been also reported for the body wall *C. lacteus* mini-Hb [21]. The finding of one oxygenated CerHb molecule per asymmetric unit in the crystalline state, together with considerations from the analysis of crystal packing contacts (which occur at scattered protein surface regions), is in keeping with the proposed O_2-linked association-dissociation equilibria, which indicate a monomeric state for the oxygenated derivative of CerHb [21].

The crystal structure of CerHb allows us to identify a new structural subgroup, henceforth called the "mini-Hb" subfamily, within the Hb homology superfamily, similar to what has been recently recognized for trHbs [14]. Remarkably, however, despite the comparable globin chain sizes, mini-Hbs and trHbs display very different structural modifications of the classical three-on-three globin α-helical sandwich, indicating diverse evolutionary histories. The structural modifications observed in the two distinct Hb subfamilies can hardly be coded in a trimmed Mb construct comprising the vertebrate Hb central exon region (topological sites B12–G7), since proper hosting of the rigid porphyrin ring within a globin matrix lacking the A and significant parts of the B and H helices requires several concerted structural adaptations, which are complex and necessarily distributed throughout the tertiary structure. Such concepts had been anticipated by characterization of a proteolyzed form of horse heart Mb, comprising residues 32–139 (topological sites B13–H16). Surprisingly, despite maintaining essential globin fold structure around the heme cavity, the protein carbonylated form is structurally closer to the parent horse Mb (60% α-helical content) than the deoxygenated form (40% α helix). Moreover, while ligand binding kinetics are similar to those of native horse Mb, the stability of the oxygenated derivative is drastically reduced [49–51]. Finally, in relation to the α-helical composition of mini-Hbs and trHbs, where α helices recognized crucial for the stability of the molten globule apoprotein intermediate are deleted or largely shortened [52–54], it may be expected that CerHb and trHb subfamilies be characterized by folding pathways markedly different from those postulated for native Mb.

Biological Implications

The smallest known Hbs (109 residues long) have been isolated from the nerve tissue and body wall muscle of

the marine nemertean worm *C. lacteus*. The nerve tissue component (CerHb) is believed to act as an oxygen store, sustaining neural activity during anoxic (burrowing) periods. The gene coding for CerHb displays two introns at the B12.2 and G7.0 sites, that is, at intron sites strongly conserved in the (non)vertebrate classical globin genes but entirely different from those displayed by truncated Hb genes. Such gene structure may place CerHb evolutionarily closer to full-length (non)vertebrate globins.

Oxygenated CerHb crystal structure shows striking adaptation of the globin fold, likely necessary to sustain heme binding in a polypeptide chain 30–50 residues shorter than conventional Hbs and Mbs. Major globin fold modifications include full deletion of the N-terminal A helix, substantial reshaping of the GH interhelical region, and reshaping of the C-terminal H helix. Poor tertiary structure matches are also found relative to the fold of truncated Hbs, the only Hb subfamily displaying polypeptide chains of a size comparable to CerHb.

Furthermore, the CerHb three-dimensional structure suggests that binding of the heme within a very short polypeptide chain requires extensive interactions with both distal and proximal residues. Clustering of heme-stabilizing distal residues may impair diffusion of O_2 to the heme Fe through the "classical" E7 gate path observed in vertebrate globins. Nevertheless, binding of O_2 to CerHb heme is supported by the Tyr(11)B10-Gln(44)E7 residue pair, through a hydrogen-bonded network reminiscent of those observed in nonvertebrate Hbs and Mbs displaying the same distal residue pair, but very different O_2 affinity.

Besides evident adaptations of the globin fold, CerHb displays a wide core hydrophobic tunnel, spanning the protein matrix from a surface cleft to the heme-ligand binding site, unique among known globin structures. This observation suggests that, in a small globin, imperfect packing of core residues around the porphyrin ring may originate tunnels which, during evolution, can acquire a functional role in supporting specific ligand (e.g., O_2) diffusion pathways to the heme. Alternatively, sizeable protein matrix tunnels may have been present in ancestral globins, and may be currently used for ligand transient storage. Such specific functional property is coded, although not as extensively as in CerHb and truncated Hbs, in Mb tertiary structure, and identified by Xe binding cavities. Thus, swinging of the E7 residue granting ligand access to the distal heme cavity of vertebrate globins may be one of several alternative mechanisms for conveying a diatomic ligand to the heme, possibly an evolutionary implementation more recent than those based on packing defects, tunnels, or cavities through the protein matrix, as independently developed by truncated Hbs and CerHb.

Experimental Procedures

Expression, Purification, and Crystallization of CerHb

A synthetic *C. lacteus* cDNA was constructed from 15 oligonucleotides using the codon frequency of *E. coli* according to the method of Ikehara et al. [55]. The gene was expressed in *E. coli* and purified as previously described [56]. Recombinant CerHb was crystallized by vapor diffusion techniques at a concentration of 27 mg/ml against a reservoir solution containing 60% ammonium sulfate and 50 mM sodium acetate buffer (pH 5.5) at 4°C. Bunches of elongated prismatic crystals ($\sim$0.07 $\times$ 0.07 $\times$ 0.4 mm^3) grew within 1 week, as previously reported [29]. The crystals were stored in 70% ammonium sulfate and 50 mM sodium acetate buffer (pH 5.5), and transferred to the same solution supplemented with 15% glycerol immediately before data collection at 100 K.

Data Collection and Processing

Three-wavelength MAD data sets were collected at ESRF synchrotron source (beamline ID29, Grenoble, France) at 100 K (Table 1). The peak and the inflection point wavelengths were determined by collecting an X-ray absorption spectrum near the heme Fe atom K absorption edge. Diffraction data were processed using DENZO, SCALEPACK, and programs from the CCP4 suite [57, 58]. The crystals belong to the orthorhombic space group $P2_12_12_1$ and accommodate one CerHb molecule per asymmetric unit, with an estimated solvent content of 47% (Table 1A).

Structure Solution and Refinement

MAD phases, based on the heme Fe atom anomalous signal, were determined at 1.95 Å resolution using SOLVE [59], with a figure of merit of 0.56. The electron density map was remarkably improved by solvent flattening using the program RESOLVE [60], yielding a figure of merit of 0.73 at 1.95 Å resolution. The resulting electron density map was of excellent quality, clearly displaying almost all the main molecular features and residues. The program wARP [61] was used to extend and refine phases to 1.5 Å resolution and for automated model building of all the main and side chain atoms. The molecular model was subsequently manually checked with O [62] and refined at the maximum resolution using CNS [63] and subsequently REFMAC [64] for anisotropic B factor refinement. The final model contains 109 residues (plus an extra N-terminal Met residue), 106 water molecules, one dioxygen molecule, and one sulfate and one acetate anions (R factor = 15.3% and R_{free} = 18.7%, respectively), with ideal stereochemical parameters (Table 1B [30]). Weak electron density was observed only for residue Lys(36)D6.

Acknowledgments

We acknowledge the European Synchrotron Radiation Facility (Grenoble, France) for provision of synchrotron radiation facilities, and wish to thank Dr. A. Thompson for support during data collection and processing at beamline ID29. This work was supported by grants from the Italian Ministry of University and Research (Project/Grant MIUR "L.95/95"), from the National Research Council of Italy ("Biotecnologie" and "Genetica Molecolare") to M.B. L.M. is supported by a grant (project number G.0069.98) of the FWO (Fund for Scientific Research—Flanders). S.D. is a postdoctoral fellow of the same fund. The EU project "Neuroglobin" QLRT-2001-01548 is fully acknowledged.

Received: December 28, 2001
Revised: February 26, 2002
Accepted: March 5, 2002

References

1. Perutz, M.F. (1979). Regulation of oxygen affinity of hemoglobin: influence of structure of the globin on the heme iron. Annu. Rev. Biochem. *48*, 327–386.
2. Wittenberg, J.B., and Wittenberg, B.A. (1989). Mechanisms of cytoplasmic hemoglobin and myoglobin function. Annu. Rev. Biophys. Biophys. Chem. *19*, 217–241.
3. Wittenberg, J.B. (1992). Functions of cytoplasmatic hemoglobins and myohemerythrin. Adv. Comp. Environ. Physiol *13*, 60–85.
4. Poole, R.K., and Hughes, M.N. (2000). New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. Mol. Microbiol. *36*, 775–783.
5. Ascenzi, P., Salvati, L., and Brunori, M. (2001). Does myoglobin protect *Trypanosoma cruzi* from the antiparasitic effects of nitric oxide? FEBS Lett. *501*, 103–105.
6. Brunori, M., and Gibson, Q.H. (2001). Cavities and packing de-

fects in the structural dynamics of myoglobin. EMBO Rep. *2*, 674–679.

7. Weber, R.E., and Vinogradov, S.N. (2001). Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol. Rev. *81*, 569–628.
8. Dickerson, R.E., and Geis, I. (1983). Hemoglobin structure and function. In Hemoglobin: Structure, Function, Evolution and Pathology (Menlo Park, CA: Benjamin/Cummings Publishing), pp. 20–64.
9. Holm, L., and Sander, C. (1993). Structural alignment of globins, phycocyanins and colicin A. FEBS Lett. *315*, 301–306.
10. Bolognesi, M., Bordo, D., Rizzi, M., Tarricone, C., and Ascenzi, P. (1997). Nonvertebrate hemoglobins: structural bases for reactivity. Prog. Biophys. Mol. Biol. *68*, 29–68.
11. Paoli, M., and Nagai, K. (2001). Hemoglobin. In Handbook of Metalloproteins, A. Messerschmidt, R. Huber, T.L. Poulos, and K. Wieghardt, eds. (Chichester, Sussex, UK: John Wiley & Sons), pp. 16–30.
12. Flögel, U., Merx, M., Godecke, A., Decking, U.K., and Schrader, J. (2001). Myoglobin: a scavenger of bioactive NO. Proc. Natl. Acad. Sci. USA *98*, 735–740.
13. Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L., and Bolognesi, M. (2000). A novel two-over-two α-helical sandwich fold is characteristic of the truncated hemoglobin family. EMBO J. *19*, 2424–2434.
14. Wittenberg, J.B., Bolognesi, M., Wittenberg, B.A., and Guertin, M. (2002). Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes and plants. J. Biol. Chem. *277*, 871–874.
15. Burmester, T., Weich, B., Reinhardt, S., and Hankeln, T. (2000). A vertebrate globin expressed in the brain. Nature *407*, 520–523.
16. Burmester, T., Ebner, B., Weich, B., and Hankeln, T. (2001). Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol. Biol. Evol. *19*, 416–421.
17. Awenius, C., Hankeln, T., and Burmester, T. (2001). Neuroglobins from the zebrafish *Danio rerio* and the pufferfish *Tetraodon nigroviridis*. Biochem. Biophys. Res. Commun. *287*, 418–421.
18. Couture, M., Burmester, T., Hankeln, T., and Rousseau, D. (2001). The heme environment of mouse neuroglobin. Evidence for the presence of two conformations of the heme pocket. J. Biol. Chem. *276*, 36377–36382.
19. Dewilde, S., Kiger, L., Burmester, T., Hankeln, T., Baudin-Creuza, V., Aerts, T., Marden, M.C., Caubergs, R., and Moens, L. (2001). Biochemical characterization and ligand-binding properties of neuroglobin, a novel member of the globin family. J. Biol. Chem. *276*, 38949–38955.
20. Dewilde, S., Blaxter, M., Van Hauwaert, M.L., Vanfleteren, J., Esmans, E.L., Marden, M., Griffon, N., and Moens, L. (1996). Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*. J. Biol. Chem. *271*, 19865–19870.
21. Vandergon, T.L., Riggs, C.K., Gorr, T.A., Colacino, J.M., and Riggs, A.F. (1998). The mini-hemoglobins in neural and body wall tissue of the nemertean worm, *Cerebratulus lacteus*. J. Biol. Chem. *273*, 16998–17011.
22. Neuwald, A.F., Liu, J.S., Lipman, D.J., and Lawrence, C.E. (1997). Extracting protein alignment models from the sequence database. Nucleic Acids Res. *25*, 1665–1677.
23. Kraus, D.W., and Colacino, J.M. (1986). Extended oxygen delivery from the nerve hemoglobin of *Tellina alternata* (Bivalvia). Science *232*, 90–92.
24. Kraus, D.W., and Doeller, J.E. (1988). A physiological comparison of bivalve mollusc cerebro-visceral connectives with and without neurohemoglobin. III. Oxygen demand. Biol. Bull. *174*, 67–76.
25. Bashford, D., Chothia, C., and Lesk, A.M. (1987). Determinants of a protein fold. Unique features of the globin amino acid sequences. J. Mol. Biol. *196*, 199–216.
26. Moens, L., Vanfleteren, J., Van de Peer, Y., Peeters, K., Kapp, O., Czeluzniak, J., Goodman, M., Blaxter, M., and Vinogradov, S. (1996). Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationship. Mol. Biol. Evol. *13*, 324–333.
27. Lesk, A.M., and Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J. Mol. Biol. *136*, 225–270.
28. Milani, M., Pesce, A., Ouellet, Y., Ascenzi, P., Guertin, M., and Bolognesi, M. (2001). *Mycobacterium tuberculosis* hemoglobin-N displays a protein tunnel suited for O_2 diffusion to the heme. EMBO J. *20*, 3902–3909.
29. Pesce, A., Nardini, M., Dewilde, S., Ascenzi, P., Riggs, A.F., Yamauchi, K., Geuens, E., Moens, L., and Bolognesi, M. (2001). Crystallization and preliminary X-ray analysis of neural hemoglobin from the nemertean worm *Cerebratulus lacteus*. Acta Crystallogr. *D57*, 1897–1899.
30. Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. *A47*, 392–400.
31. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK, a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
32. Pesce, A., Dewilde, S., Kiger, L., Milani, M., Ascenzi, P., Marden, M.C., Van Hauwaert, M.L., Vanfleteren, J., Moens, L., and Bolognesi, M. (2001). Very high resolution structure of a trematode hemoglobin displaying a TyrB10-TyrE7 heme distal residue pair and high oxygen affinity. J. Mol. Biol. *309*, 1153–1164.
33. Tarricone, C., Galizzi, A., Coda, A., Ascenzi, P., and Bolognesi, M. (1997). Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla* sp. Structure *5*, 497–507.
34. Kapp, O.H., Moens, L., Vanfleteren, J., Trotman, C.N., Suzuki, T., and Vinogradov, S.N. (1995). Alignment of 700 globin sequences: extent of amino acid substitution and its correlation with variation in volume. Protein Sci. *10*, 2179–2190.
35. Rizzi, M., Wittenberg, J.B., Coda, A., Fasano, M., Ascenzi, P., and Bolognesi, M. (1994). Structure of the sulfide-reactive hemoglobin from the clam *Lucina pectinata*. Crystallographic analysis at 1.5 Å resolution. J. Mol. Biol. *244*, 86–99.
36. Bisig, D.A., Di Iorio, E.E., Diederichs, K., Winterhalter, K.H., and Piontek, K. (1995). Crystal structure of Asian elephant (*Elephas maximus*) cyano-metmyoglobin at 1.78-Å resolution. Phe29(B10) accounts for its unusual ligand binding properties. J. Biol. Chem. *270*, 20754–20762.
37. Minning, D.M., Gow, A.J., Bonaventura, J., Braun, R., Dewhirst, M., Goldberg, D.E., and Stamler, J.S. (1999). *Ascaris* haemoglobin is a nitric oxide-activated 'deoxygenase'. Nature *401*, 497–502.
38. Burley, S.K., and Petsko, G.A. (1988). Weakly polar interactions in proteins. Adv. Protein Chem. *39*, 125–189.
39. Bolognesi, M., Rosano, C., Losso, R., Borassi, A., Rizzi, M., Wittenberg, J.B., Boffi, A., and Ascenzi, P. (1999). Cyanide binding to *Lucina pectinata* hemoglobin I and to sperm whale myoglobin: an X-ray crystallographic study. Biophys. J. *77*, 1093–1099.
40. Hubbard, S.J., Gross, K.H., and Argos, P. (1994). Intramolecular cavities in globular proteins. Protein Eng. *7*, 613–626.
41. Perutz, M.F. (1989). Myoglobin and haemoglobin: role of distal residues in reactions with haem ligands. Trends Biochem. Sci. *14*, 42–44.
42. Tilton, R.F., Jr., Kuntz, I.D., Jr., and Petsko, G.A. (1984). Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 Å. Biochemistry *23*, 2849–2857.
43. Brunori, M. (2001). Nitric oxide, cytochrome-c oxidase and myoglobin. Trends Biochem. Sci. *26*, 21–23.
44. Brunori, M. (2001). Nitric oxide moves myoglobin centre stage. Trends Biochem. Sci. *26*, 209–210.
45. Brunori, M., Vallone, B., Cutruzzolá, F., Travaglini-Allocatelli, C., Berendzen, J., Chu, K., Sweet, R.M., and Schlichting, I. (2001). The role of cavities in protein dynamics: crystal structure of a photolytic intermediate of a mutant myoglobin. Proc. Natl. Acad. Sci. USA *97*, 2058–2063.
46. Frauenfelder, H., McMahon, B.H., Austin, R.H., Chu, K., and Groves, J.T. (2001). The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. Proc. Natl. Acad. Sci. USA *98*, 2370–2374.
47. Scott, E.E., Gibson, Q.H., and Olson, J.S. (2001). Mapping the

pathway for O_2 entry into and exit from myoglobin. J. Biol. Chem. *276*, 5177–5188.
48. Royer, W.E., Jr., Knapp, J.E., Strand, K., and Heaslet, H.A. (2001). Cooperative hemoglobins: conserved fold, diverse quaternary assemblies and allosteric mechanisms. Trends Biochem. Sci. *26*, 297–304.
49. De Sanctis, G., Falcioni, G., Giardina, B., Ascoli, F., and Brunori, M. (1988). Mini-myoglobin. The structural significance of haem-ligand interactions. J. Mol. Biol. *200*, 725–733.
50. De Sanctis, G., Falcioni, G., Grelloni, F., Desideri, A., Polizio, F., Giardina, B., Ascoli, F., and Brunori, M. (1991). Mini-myoglobin. Electron paramagnetic resonance and reversible oxygenation of the cobalt derivative. J. Mol. Biol. *222*, 637–643.
51. De Sanctis, G., Ascoli, F., and Brunori, M. (1994). Folding of apominimyoglobin. Proc. Natl. Acad. Sci. USA *91*, 11507–11511.
52. Barrick, D., and Baldwin, R.L. (1993). Stein and Moore award address. The molten globule intermediate of apomyoglobin and the process of protein folding. Protein Sci. *2*, 869–876.
53. Jennings, P.A., and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science *262*, 892–896.
54. Tcherkasskaya, O., and Uversky, V.N. (2001). Denatured collapsed states in protein folding: example of apomyoglobin. Proteins *44*, 244–254.
55. Ikehara, M., Ohtsuka, E., Tokunaga, T., Taniyama, Y., Iwai, S., Kitano, K., Miyamoto, S., Ohgi, T., Sakuragawa, Y., Fujiyama, K., et al. (1984). Synthesis of a gene for human growth hormone and its expression in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *81*, 5956–5960.
56. Dewilde, S., Blaxter, M., Van Hauwaert, M.L., Van Houte, K., Pesce, A., Griffon, N., Kiger, L., Marden, M.C., Vermeire, S., Vanfleteren, J., et al. (1998). Structural, functional and genetic characterization of *Gastrophilus* hemoglobin. J. Biol. Chem. *273*, 32467–32474.
57. CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. *D50*, 760–763.
58. Otwinoski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
59. Terwilliger, T.C., and Berendzen, J. (1999). Automated structure solution for MIR and MAD. Acta Crystallogr. *D55*, 849–861.
60. Terwilliger, T.C. (2000). Maximum likelihood density modification. Acta Crystallogr. *D56*, 965–972.
61. Perrakis, A., Morris, R., and Lamzin, V. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.
62. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.
63. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905–921.
64. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. *D53*, 240–255.
65. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.
66. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
67. Merritt, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.
68. Laskowski, R.A. (1995). SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. J. Mol. Graph. *13*, 323–330.
69. Esnouf, B.M. (1997). An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 138.

Accession Numbers

CerHb coordinates and structure factors have been deposited in the Protein Data Bank under code 1KR7 [65].