The EMBO Journal Vol. 19 No. 11 pp. 2424–2434, 2000

A novel two-over-two α-helical sandwich fold is characteristic of the truncated hemoglobin family

Alessandra Pesce[1], Manon Couture[2,3], Sylvia Dewilde[4], Michel Guertin[2], Kiyoshi Yamauchi[5], Paolo Ascenzi[1,6], Luc Moens[4] and Martino Bolognesi[1,7]

[1]Department of Physics – INFM and Advanced Biotechnology Center – IST, University of Genova, Largo Rosanna Benzi 10, 16132 Genova, [6]Department of Biology, University of 'Roma Tre', Viale Guglielmo Marconi 446, 00146 Roma, Italy, [2]Departement de Biochimie et de Microbiologie, Pavillon Marchand, Universitè Laval, Quebec, G1K 7P4, Canada, [4]Department of Biochemistry, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium and [5]Department of Biology and Geoscience, Faculty of Science, Shizuoka University, 836 Oya, Shizuoka 422-8529, Japan

[3]Present address: Department of Physiology and Biophysics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, The Bronx, NY 10461, USA

[7]Corresponding author
e-mail: bolognes@fisica.unige.it

Small hemoproteins displaying amino acid sequences 20–40 residues shorter than (non-)vertebrate hemoglobins (Hbs) have recently been identified in several pathogenic and non-pathogenic unicellular organisms, and named 'truncated hemoglobins' (trHbs). They have been proposed to be involved not only in oxygen transport but also in other biological functions, such as protection against reactive nitrogen species, photosynthesis or to act as terminal oxidases. Crystal structures of trHbs from the ciliated protozoan *Paramecium caudatum* and the green unicellular alga *Chlamydomonas eugametos* show that the tertiary structure of both proteins is based on a 'two-over-two' α-helical sandwich, reflecting an unprecedented editing of the classical 'three-over-three' α-helical globin fold. Based on specific Gly–Gly motifs the tertiary structure accommodates the deletion of the N-terminal A-helix and replacement of the crucial heme-binding F-helix with an extended polypeptide loop. Additionally, concerted structural modifications allow burying of the heme group and define the distal site, which hosts a TyrB10, GlnE7 residue pair. A set of structural and amino acid sequence consensus rules for stabilizing the fold and the bound heme in the trHbs homology subfamily is deduced.

Keywords: chloroplast hemoglobin/globin-fold evolution/helical fold/hemoglobins/*Paramecium* hemoglobin

Introduction

The discovery of hemoglobins (Hbs) in virtually all kingdoms has shown that the gene for Hb is very ancient and that Hbs may serve functions other than as simple oxygen carriers (Hardison, 1998; Imai, 1999; Minning *et al*., 1999). Within unicellular organisms, two different Hbs, or Hb-related protein groups, are found. A first group, occurring in bacteria and fungi, includes single-chain flavohemoglobins, which consist of an N-terminal heme-containing domain displaying a conventional globin fold, and a C-terminal FAD or $NADP^+$ binding domain structurally related to ferredoxin $NADP^+$ reductase (Ermler *et al*., 1995). Moreover, a dimeric Hb, which may be non-covalently associated to a flavoprotein reductase, has been isolated from *Vitreoscilla* sp. The three-dimensional structure of the N-terminal heme domain of flavohemoglobin and of *Vitreoscilla* Hb conforms to the classical vertebrate and non-vertebrate Hb fold, based on seven or eight α-helices (Perutz, 1979; Holm and Sander, 1993); however, the homodimeric *Vitreoscilla* Hb adopts a quaternary structure not observed in any other known Hb (Bolognesi *et al*., 1997, 1999a; Tarricone *et al*., 1997).

The second protein group includes small hemoproteins, refered to as truncated hemoglobins (trHbs), characterized so far in the ciliated protozoa *Paramecium caudatum* and *Tetrahymena pyriformis*, in the unicellular alga *Chlamydomonas eugametos*, and in the eubacteria *Nostoc commune* and *Mycobacterium tuberculosis* (Iwaasa *et al*., 1989; Potts *et al*., 1992; Takagi, 1993; Couture *et al*., 1994, 1999a,b; Thorsteinsson *et al*., 1999). To these should be added those being discovered through the sequencing of microbial genomes (see Figure 1). In this regard, analysis of the currently available microbial genomic sequences indicates that certain eubacteria (*Bacillus subtilis*, *Staphylococcus aureus*, *Campylobacter jejuni*, *Bordetella pertussis* and *Deinococcus radiodurans*) contain both flavohemoglobin and trHb, which suggests specific functions for each of the two protein classes.

TrHbs of 110–130 amino acids per protein chain (Figure 1) have been isolated as monomeric and dimeric species, displaying medium to extreme oxygen affinities, with a case of very high ligand binding cooperativity (Couture *et al*., 1999a,b). They all show very low amino acid sequence homology to vertebrate and non-vertebrate Hbs, with sequence identities <15% (Couture *et al*., 1994; Moens *et al*., 1996). Sequence alignments reveal substantial residue deletions at either N- or C-termini and in the CD–D region of the (non-)vertebrate globin fold. [α-helices building up the globin fold in vertebrate Hbs are conventionally labeled A, B…H according to their sequential order; topological sites are numbered sequentially within each α-helix (Perutz, 1979)]. The distal residue at the E7 position, which is almost invariably His or Gln in (non-)vertebrate Hbs, and which stabilizes the heme-bound ligand through hydrogen bonding, is often Gln in trHbs. However, residues not capable of hydrogen bonding may also be found at the E7 position (Figure 1). The other distal residue, at position B10, is invariably Tyr

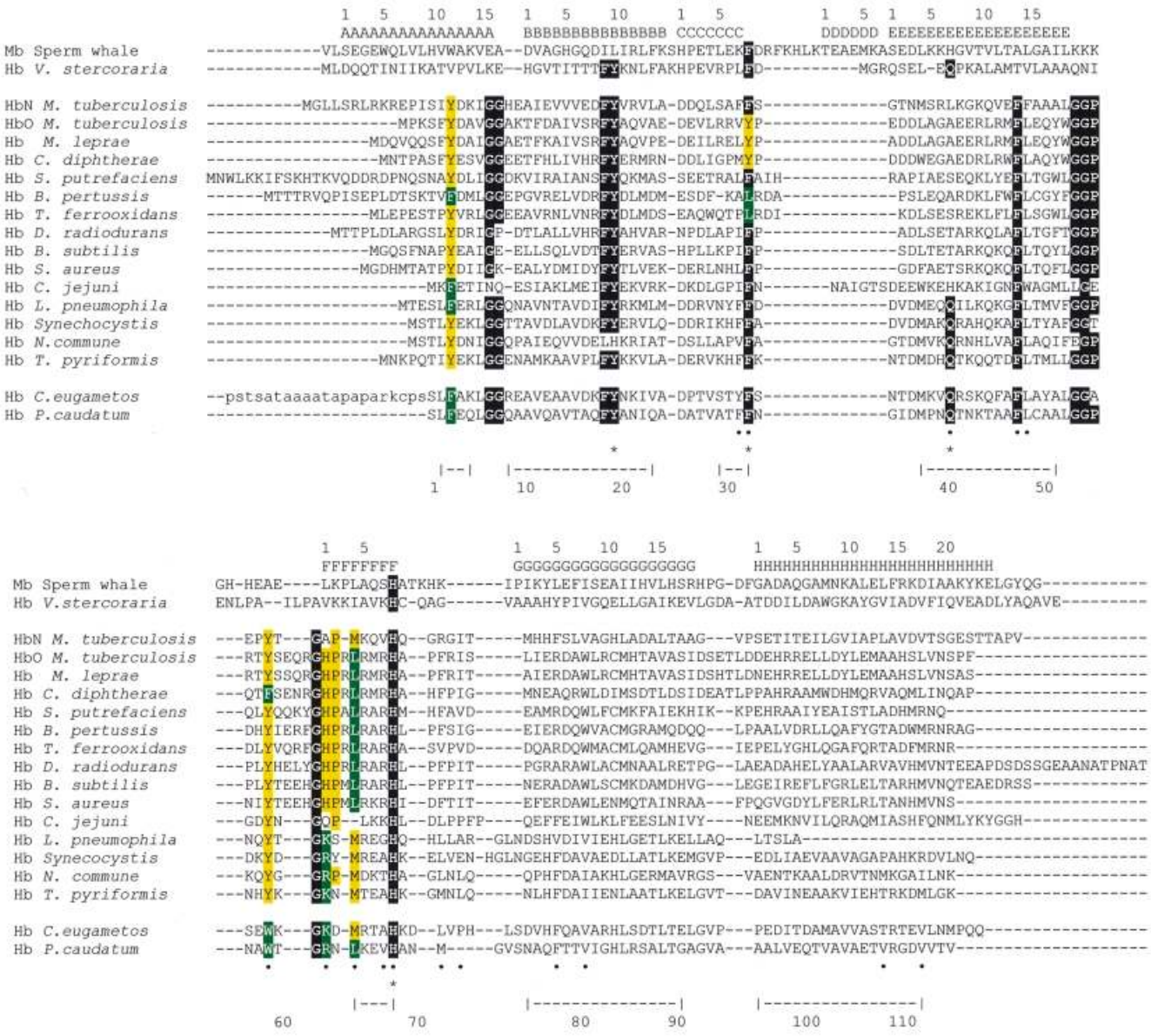

Fig. 1. Structure-based sequence alignment of PtrHb, CtrHb and 15 trHbs from different sources with respect to sperm whale (*Physeter catodon*) Mb and *Vitreoscilla* sp. Hb, the latter taken as reference vertebrate and bacterial globin molecules, respectively. The globin fold topological positions, as defined in sperm whale Mb, are shown on the top of the aligned sequences. Amino acid sequential numbering, as well as α-helical regions (indicated by |--| segments), refer to PtrHb. Heme-contacting residues are indicated by a bullet, and some key globin fold topological sites by a star. Residues that are conserved in all trHbs are highlighted in black boxes; yellow and green bars highlight amino acid sites, with conservative substitutions in trHbs, which are discussed in the text in relation to the achievement of the trHb globin fold. Sequence DDBJ/EMBL/GenBank accession numbers are: *M.tuberculosis* HbN, Z74020; *M.tuberculosis* HbO, AL021246; *B.subtilis* Hb, Z99110; *N.commune* Hb, L47979; *Synechocystis* PCC6803 Hb, D90910; *P.caudatum* Hb, M57542; *T.pyriformis* Hb, D13920; *C.eugametos* Hb, X72919. Preliminary sequence data for *S.aureus*, *Shewanella putrefaciens*, *D.radiodurans* and *T.ferrooxidans* Hbs were obtained from the Institute for Genomic Research website at http:/www.tigr.org. Sequence data for *B.pertussis*, *C.diphtheriae*, *C.jejuni* and *M.leprae* were produced by the Sequencing Group at the Sanger Center and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/cj and ftp://ftp.sanger.ac.uk/pub/pathogens/bp, respectively. Sequence data for *Legionella pneumophila* were produced by the Columbia Genome Center, and can be obtained from the 'Unfinished Microbial Genomes Database at NCBI'.

in trHbs. Its role in affecting ligand stability by complementing the hydrogen bonding capability of the E7 residue has been shown in non-vertebrate Hbs (Yang *et al*., 1995) and in the trHbs from *C.eugametos* and from *M.tuberculosis* (Couture *et al*., 1999a,b). *Nostoc commune* trHb, with a His residue at position B10, is the only exception. The structural implications of such a remarkable evolutionary divergence, with respect to the wide and highly characterized family of (non-)vertebrate Hbs, are unknown since no trHb three-dimensional structure has been determined so far. On the other hand, a recent NMR investigation on cyano-met *N.commune* trHb, focused on the heme crevice structural features, has suggested conservation of the F-, G- and H-helices and of the FG hinge region (Yeh *et al*., 2000).

The functional roles of trHbs are virtually unknown and may be various. *Nostoc commune* trHb may be a component of a microaerobically induced terminal oxidase (Potts *et al*., 1992; Thorsteinsson *et al*., 1999). In the unicellular green alga *C.eugametos* a trHb is induced in

Table I. Data collection and refinement statistics for PtrHb and CtrHb

(A) PtrHb MAD data collection statistics

	Absorption peak	Inflection point	Remote
Wavelength (Å)	1.739	1.740	0.979
Resolution (Å)	30–2.7	30–2.7	30–1.54
Mosaicity (°)	0.31	0.31	0.31
Completeness (%)	97.6 (91.2)[a]	97.2 (86.6)	96.4 (88.2)
R_{merge} (%)	2.2	2.2	2.6
Independent reflections	3681	3723	19397
Average $I/\sigma(I)$	32 (13)	32 (13)	31 (7)

(B) CtrHb MIR data collection statistics

	Native	$UO_2(Ac)_2$	K_2PtCl_4
Wavelength (Å)	1.00	1.542	1.542
Resolution (Å)	30–1.80	16–3.1	16–3.1
Mosaicity (°)	0.47	0.24	0.76
Completeness (%)	98.6 (99.6)	98.7 (71.3)	95.8 (68.5)
R_{merge} (%)	5.6	21.7	19.2
R_{deriv} (%)	–	37.0	31.1
Independent reflections	12839	2508	2358
Average $I/\sigma(I)$	21 (10)	9 (3)	8 (3)
Phasing power[b]	–	1.59	1.18
Number of sites	–	1	2

(C) Refinement statistics and model quality

	PtrHb	CtrHb
Resolution range (Å)	30–1.54	30–1.80
Total number of non-hydrogen atoms	881	978
Number of water molecules	207	186
R-factor	0.133	0.176
R_{free}[c]	0.183	0.211
Space group	$P4_3$	$P2_12_12_1$
Unit cell (Å)	$a = 61.2$	$a = 34.6$
	$b = 61.2$	$b = 53.1$
	$c = 35.8$	$c = 67.2$
R.m.s.d. from ideal geometry		
bond lengths (Å)	0.014	0.018
bond angles (°)	1.57	1.65
Ramachandran plot[d]		
most favoured region	96%	99%
additional allowed region	4%	1%
Averaged *B*-factors (Å^2)		
main chain	11	15
side chain	14	17
solvent	27	31
heme	13	13

[a]Outer shell statistics are shown in parentheses. The outer shells are 2.75–2.70 Å for the absorption peak and inflection point, and 1.58–1.54 Å for remote point in PtrHb; and 1.86–1.80 Å for native and 3.15–3.10 Å for heavy atom derivatives in CtrHb.
[b]Phasing power: r.m.s. ($|F_H|/E$), where $|F_H|$ is the heavy atom structure factor amplitude and E is the lack of closure ($|F_{PH} - F_P| - |F_H|$).
[c]Calculated using 10% of the reflections.
[d]Data produced using the program PROCHECK (Laskowski *et al.*, 1993).

response to activated photosynthesis and is localized, in part, along the thylakoid membranes (Couture *et al.*, 1994). In *M.tuberculosis*, a trHb has been postulated to be involved in the protection of the bacilli against reactive nitrogen species produced by the host (Couture *et al.*, 1999b).

We report here the results of an X-ray crystallographic investigation on the trHbs isolated from the protist *P.caudatum* (PtrHb; 116 amino acids, in the aquo-met form) and from the green unicellular alga *C.eugametos* (CtrHb; 121 amino acids, in the cyano-met form), whose three-dimensional structures were determined independently. The emphasis of the discussion is on the unique modifications that are built onto the globin fold of both trHbs, on their heme distal binding sites, and on the structural requirements that allow heme stabilization within a very short, possibly minimal, Hb-like polypeptide chain.

Results and discussion

Structure determination

The crystal structure of PtrHb was solved by means of multiple wavelength anomalous diffraction (MAD) based on the anomalous scattering of the heme Fe atom. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at 100 K. The final protein model was refined

at 1.54 Å resolution, with close to ideal stereochemical parameters, to an *R*-factor of 13.3% (R_{free} = 18.3%; see Table I). The electron density accounts for all 116 residues of recombinant PtrHb, and for 207 ordered water molecules. The CtrHb three-dimensional structure was solved by means of multiple isomorphous replacement, based on two heavy atom derivatives (see Table I), using a rotating anode X-ray source at room temperature. The structure was refined at 1.8 Å resolution, using a native data set collected at ESRF, at 100 K, to an *R*-factor of 17.6% (R_{free} = 21.1%) with good overall stereochemistry. The protein model accounts for all 121 residues of the expressed protein, including 186 water molecules.

The trHb fold

PtrHb and CtrHb tertiary structures are very similar (the r.m.s.d. calculated over 107 C_α atom pairs is 0.97 Å), and display a molecular fold based on four main α-helices (see Figure 2). The α-helices around the heme group are arranged in a sort of bundle, composed of two antiparallel helix pairs (labeled B/E and G/H in Figure 2A), connected by an extended polypeptide loop. If attention is focused on the protein chain only, some structural resemblance to the ferritin subunit fold may be envisaged (Hempstead *et al*., 1997). Nevertheless, the secondary structure elements are arranged in different topological order in the two protein families. More properly, the trHb fold can be recognized as a subset of the classical 'three-over-three' α-helical sandwich (Holm and Sander, 1993), trimmed to a domain essentially composed of helices B, E, G and H, in a 'two-over-two' arrangement. Structural superpositions of PtrHb and CtrHb with sperm whale myoglobin (Mb; limited to the common α-helical regions, 59 C_α atom pairs) yield r.m.s.d. values in the 2.3 Å range (Figure 2B). On the other hand, superpositions with *Vitreoscilla* sp. Hb, the only other bacterial Hb structure known (Tarricone *et al*., 1997; Bolognesi *et al*., 1999a) (r.m.s.d. values of ~2.8 Å; Figure 2C), are limited to 44 C_α pairs only, due to different α-helix orientations. Both sets of comparison highlight extensive modifications of the trHb fold with respect to more conventional Hb folds, particularly in the A-helix and in the CD–D regions, which are virtually absent, and in the EF–F regions (see below). Figure 2B and C show that from the tertiary structure viewpoint the protein region distal to the heme is more conserved in trHbs than the proximal half of the molecule, which appears substantially trimmed as compared with (non-)vertebrate and other bacterial Hbs.

Structural features of trHbs on the heme distal side

In both trHbs, the regular structure achieved along the B- and E-helices allows one to identify several key residues relative to the heme distal ligand binding site. Residue TyrB10, which is buried in the inner part of the heme pocket and properly oriented through hydrogen bonds to residues GlnE7 and Thr/GlnE11 in PtrHb/CtrHb, respectively (in the following, unless specified otherwise, whenever residues belonging to the same topological position in PtrHb and CtrHb, respectively, are different they are explicitly listed and separated by a solidus), provides stabilization of the heme-bound distal ligand (see Figure 3). In fact, TyrB10 is strongly hydrogen bonded to a heme-coordinated water molecule in PtrHb (2.76 Å) and to the distal N atom of the heme iron-bound cyanide in CtrHb. However, the cyanide electron density and its coordination geometry (see Table II) suggest that the diatomic ligand may be present in more than one binding mode, indicative of X-ray-induced (partial) heme Fe reduction (Bolognesi *et al*., 1999b). The interaction of TyrB10 with the distal ligand(s) bears particular functional relevance because a distal HisE7 residue is absent in all trHbs, with the only exception being the trHb identified in the *C.jejuni* genome (see Figure 1).

The GlnE7 side chain is pointing into the distal site cavity, providing one additional hydrogen bond to the heme-bound ligand in PtrHb [GlnE7 NE2–O(water) 2.73 Å] and in CtrHb [GlnE7 NE2–N(cyanide) 2.92 Å; see Figure 3]. Next, Thr/GlnE11, located one α-helical turn from the distal E7 residue and in contact with the heme, are moderately polar residues that are never found at this topological site in (non-)vertebrate Hbs, where they are invariantly of hydrophobic nature (Val, Leu, Ile) (Bashford *et al*., 1987; Kapp *et al*., 1995; Bolognesi *et al*., 1997). In both trHbs, the E11 residue is hydrogen bonded to the TyrB10 side chain (2.59 and 2.95 Å in PtrHb and CtrHb, respectively). Through minor fluctuations (<0.4 Å) from the observed conformation, residue E11 may further extend the interlaced distal site hydrogen-bonded network, which includes TyrB10, GlnE7, residue E11 and the heme-bound ligand. No other species (e.g. H_2O or acetate) is present in the two distal sites.

In both trHbs, residue LysE10 is solvent exposed, but bent towards the heme propionates, with which it is electrostatically linked. Particularly in CtrHb, LysE10 is also hydrogen bonded to the A propionate carboxylate (2.82 Å; see Figure 4). The strong conservation of a basic residue at this site may suggest a role in heme stabilization, a particularly relevant function in this Hb subfamily where the heme:protein interactions on the proximal side are substantially modified (see below). Indeed, heme dissociation from *N.commune* trHb has been evaluated as being ~100-fold faster than in sperm whale Mb (Thorsteinsson *et al*., 1999).

Residue PheE14 is close to the EF interhelical region and is almost orthogonal to the porphyrin ring, at the heme CHD methinic bridge in both trHbs (Figure 4). This arrangement provides a rigid closure to the heme pocket, making use of a bulky aromatic residue that is strictly conserved throughout the known trHb amino acid sequences (Figure 1). Such a structural feature may be dictated by the specific orientation of the E-helix, caused by the extremely short segment following the C region, and by the virtual absence of the F-helix in trHbs (Figure 2). It can be suggested that PheE14 is serving a structural role comparable to that ascribed to PheCD1 in (non-)vertebrate and bacterial Hbs, shielding the heme from the solvent at a porphyrin ring location remote and almost opposite to CD1. Furthermore, residue Trp59 (in both trHbs), in the proximal pre-F segment, is in van der Waals contact with and orthogonal to the side chain of PheE14. Conservation of an aromatic residue at sequence site 59 is a key primary structure feature in all the trHbs (Figure 1), suggesting that both this and the E14 side chains play a relevant role in preventing solvent-induced heme oxidation.

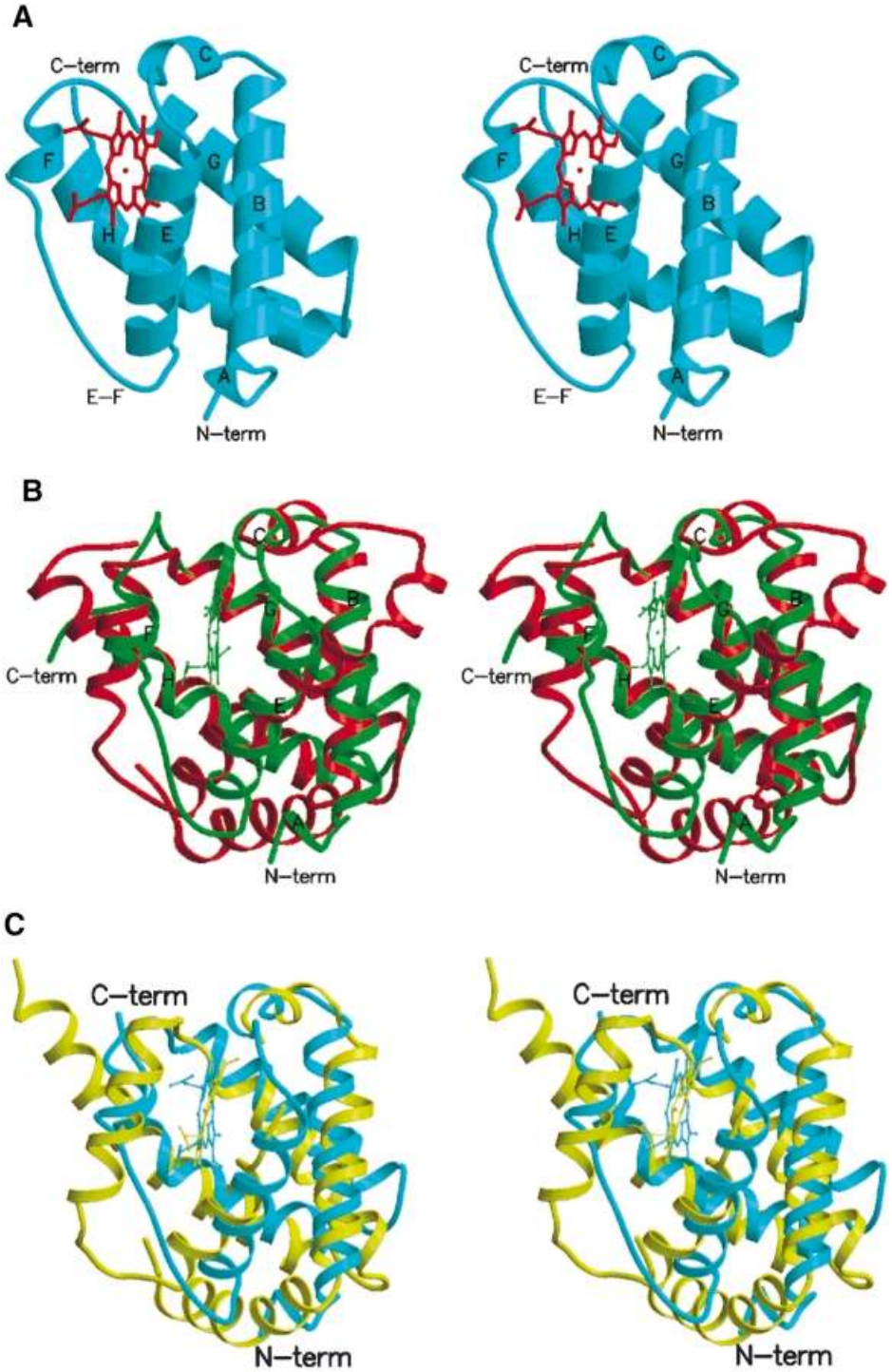

Fig. 2. (**A**) A ribbon stereo view of PtrHb tertiary structure, including the heme group. α-helices are labeled according to the conventional globin fold nomenclature (Perutz, 1979). The protein loop identified as 'pre-F' in the text immediately follows the labeled E–F region, and precedes the F one-helical turn on the heme proximal side (left in the figure). (**B**) A stereo view of the structural overlay of CtrHb (green) and sperm whale Mb (red). The protein molecules are oriented approximately as in (A). The heme group of sperm whale Mb has been omitted for clarity. (**C**) Overlay of PtrHb (light blue) and the A chain of the homodimeric *Vitreoscilla* Hb (yellow, PDB code 1vhb), in approximately the same orientation as (A) and (B). The heme group of *Vitreoscilla* Hb is also included. All figures were drawn with MOLSCRIPT (Kraulis, 1991).

Despite the fact that both PtrHb and CtrHb display TyrB10 and GlnE7 distal residues, and share a comparable distal site structure, they display very different O_2 affinities (P_{50} values are 0.49 and <0.005 mm Hg for PtrHb and CtrHb, respectively) (Couture *et al*., 1999a). This suggests that other factors may play a role in governing the overall ligand affinity of the two trHbs, such as kinetics (Couture *et al*., 1999a), or the ligand access, through protein motions, to secondary docking sites (Brunori *et al*., 1999a,b; Chu *et al*., 2000). It may be

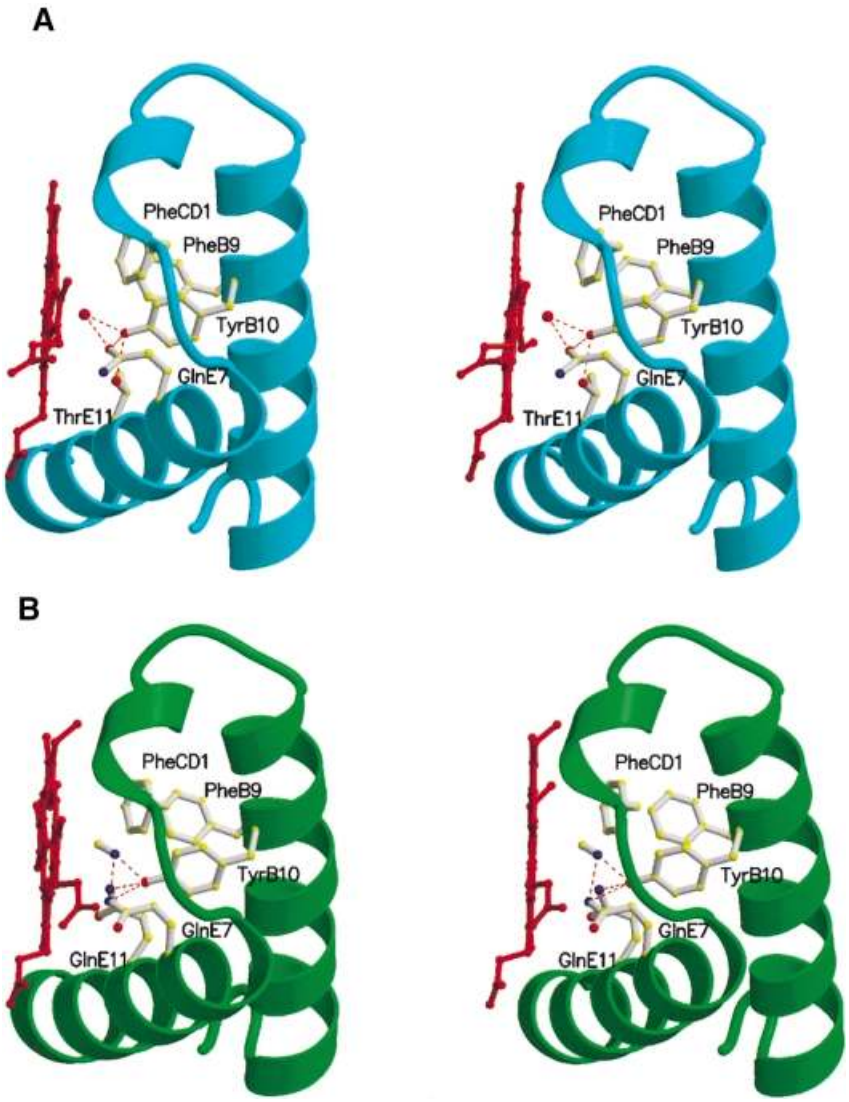

Fig. 3. Stereo view of the heme distal site main residues, together with the B-, C- and E-helices, in aquo-met PtrHb (**A**) and cyano-met CtrHb (**B**). Hydrogen bonds within the distal residues cluster, including the heme ligand, are indicated with dashed lines; for reference, the PheCD1 residue is included. In both views the heme is edge on, and the protein moiety has been partly rotated along the horizontal axis with respect to the orientation adopted for Figure 2.

noted that the hydrogen bonds and contacts within the CtrHb distal site are looser than in PtrHb despite the presence of a diatomic ligand (cyanide) bulkier than the distal water molecule of aquo-met PtrHb. This observation should be related to the known capability of ferric CtrHb to bind unusual ligands such as β-mercaptoethanol or dithiothreitol (Couture and Guertin, 1996), or to form a 6-coordinate low-spin heme complex with TyrB10, at alkaline pH (Das *et al*., 1999). Formation of the latter complex requires the TyrB10 phenolic O atom to shift by 3.5 Å with respect to its position in the cyano-met complex, indicating a remarkable distal site structure adaptability to the incoming ligand(s) in CtrHb. In this respect, engineering of distal site residues in sperm whale Mb (bearing LeuB10 in the wild-type species) has shown that the precise mutual orientation of the supporting α-helices is a primary factor in making TyrB10 available for distal ligand stabilization through direct hydrogen bonding (Brunori *et al*., 1999a).

A distal site structural organization reminiscent of that of the two trHbs has previously been observed in *Ascaris suum* Hb, which displays TyrB10 and GlnE7 hydrogen bonded to the heme coordinated dioxygen, but hosts an apolar Ile residue at the E11 site (Yang *et al*., 1995). Remarkably, *Ascaris* Hb displays a very high oxygen affinity (P_{50} = 0.004 mm Hg) and has recently been proposed to be a true enzyme, acting as an NO-activated deoxygenase (Minning *et al*., 1999).

Unique features in the trHb fold

Among the many structural features of trHbs that deviate from the conventional globin fold, the deletion of the A-helix and the presence of an extended loop substituting for most of the F-helix deserve particular consideration. Indeed, these two features may be structurally correlated. A block deletion of 11 residues in the N-terminal A-helix region (see Figures 1 and 2) might appear to threaten protein stability, since anchoring of the N-terminal region

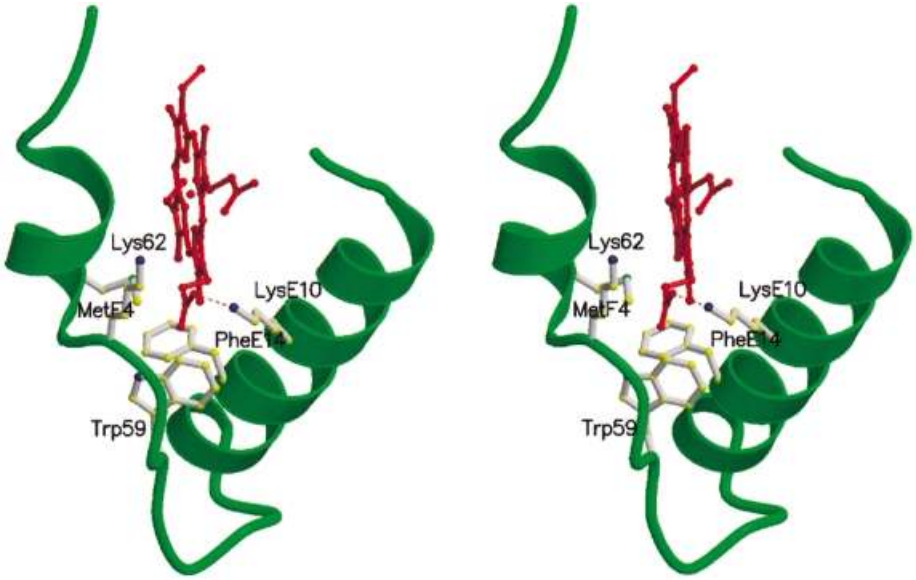

Fig. 4. A stereo view of the hydrophobic cluster preventing solvent access to the heme pocket from the pre-F region, in CtrHb. The figure includes the heme group, the E-helix, the pre-F loop and the F α-helical turn. Residues LysE10 and Lys62, which are electrostatically linked to the heme propionates, are also displayed.

Table II. Coordination geometry at the Fe(III) heme centers

Coordination bond	Distance (Å)
PtrHb	
average Fe–N(pyrrole)	2.01
HisF8 NE2 —Fe	2.14
Fe–O(ligand)	2.09
CtrHb	
average Fe–N(pyrrole)	1.98
HisF8 NE2–Fe	2.18
Fe–C(ligand)[a]	2.65
	Angle (°)
Fe–C–N[a]	130

[a]Owing to partial reduction of the heme Fe center, during the X-ray data collection stage, the cyanide coordination geometry reflects a structural average between a regularly Fe(III) coordinated ligand and Fe(II) unbound species, trapped in the distal site cavity. The *B*-factors for cyanide C and N atoms are 15 and 21 Å^2, respectively.

to the EF corner through a conserved hydrophobic contact could be impaired (Lesk and Chothia, 1980; Bashford *et al.*, 1987). Nevertheless, a compact trHb molecule with efficient sealing of the proximal side of the heme pocket is achieved through a hydrophobic cluster located in the AB topological region, previously identified in non-vertebrate globin structures (Bolognesi *et al.*, 1997). In fact, the conservation of just the last A-helix turn allows the trHb N-terminus to lock onto the protein core, through clustering of residues LeuA11, PheA12, LeuA15, Val/AlaB5, LeuE15, LeuE19 and Val/IleH4. Inspection of the aligned sequences of trHbs shows that residue A12 is always either Phe or Tyr, and that a Gly–Gly motif, coding the required structural adaptability with respect to (non-)vertebrate Hbs, is strongly conserved in the AB region (Figure 1).

In both trHbs the N-terminal dipeptide (topological sites A10–A11) runs antiparallel to the EF corner Leu(53)–Gly(54) segment, allowing hydrogen bonding between the protein backbone atoms of the two segments and further anchoring of the N-terminal region to the protein core. As a result of such antiparallel pairing, the protein backbone following the linear Leu(53)–Gly(54)–Gly(55)–Pro/Ala(56) EF peptide protrudes markedly from the protein core, giving rise to a unique proximal side structural arrangement (see Figures 2, 4 and 5). The Ramachandran φ,ψ angles adopted by residues 54–55 in both trHbs indicate the strict requirement for such a Gly–Gly sequence motif to code for their backbone conformation. The key structural role played by these residues in the trHb fold is further stressed by the conservation of the Gly–Gly–(Pro) motif in the EF region of the aligned amino acid sequences (see Figure 1).

The seven residues following the EF corner trace a wide loop structure (the 'pre-F loop'), which provides contacts with the C-terminal half of the E-helix and specific interactions with the porphyrin group. A single turn of α-helical conformation is achieved only at residues Leu/MetF4–HisF8, proximal to the heme iron atom (see Figures 4 and 5). The absence of essentially the whole F-helix on the proximal side of the heme has never been observed before in Hbs, and is reflected in PtrHb and CtrHb tertiary structures by a (proximal) wide and shallow protein surface pocket (~12 Å diameter), delimited by segments of the E-helix, of the pre-F loop, by residues F4–F5 and by part of the H-helix. The pocket hosts the bulky side chains of PheE14, Trp(59) and Leu/MetF4, which primarily support the pre-F loop structure and shield the porphyrin from solvent, on the proximal side. Indeed, in both trHbs at least six ordered solvent molecules are found hydrogen bonded within the shallow pocket.

Further support to the well defined structure of the pre-F loop, as well as to stabilization of the bound porphyrin ring, is provided in PtrHb by a strong hydrogen-bonded salt bridge, connecting Arg(62) to the heme propionate D carboxylate (2.84 Å). On the other hand, in CtrHb, residue Lys(62) is electrostatically linked to the heme A propionate. In this respect, it should be noted that a Gly–Arg/

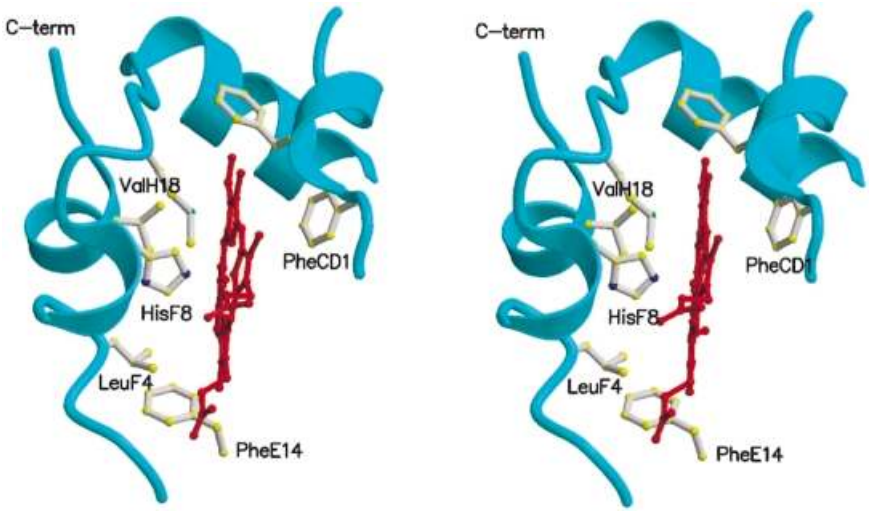

Fig. 5. Proximal side of the heme in PtrHb, displaying the one-turn F-helix, residues LeuF4 and HisF8, together with the heme group. The figure portrays part of the heme crevice, defined by the C-helix (PheC7 and PheCD1 are shown), and by segments of the G-helix and of the H-helix (ValH18 and the C-terminus are shown). For reference, residue PheE14, in the lower part of the heme pocket, is included.

Lys(or His) sequence motif is strongly conserved in the trHb amino acid sequences (the Gly residue being truly invariant), at the putative end of their pre-F loops (Figure 1), suggesting conservation of the salt bridge interaction with a heme propionate, as noted above for the comparable electrostatic interaction played by LysE10, on the distal side (see Figure 4).

Following the pre-F loop, the F α-helical single turn provides contacts between residue Leu/MetF4 and the porphyrin ring, and a hydrogen bond connecting the F4 carbonyl O atom to the proximal HisF8 ND1 atom. Both interactions are commonly present in (non-)vertebrate Hbs (Bolognesi *et al*., 1997); the latter defines the F8 imidazole ring azimuthal orientation, which is almost identical in the two trHbs and staggered with respect to the heme pyrrole N atoms. Despite the very limited span of the F-helix region, proximal coordination to the heme is regularly achieved. The iron atom is offset by 0.16 Å towards the proximal side with respect to the heme pyrrole N atoms plane in PtrHb, while it is essentially within this plane in CtrHb. Overall, the heme Fe coordination geometry is in agreement with known model compounds and with previous observations for (non-)vertebrate or bacterial Hbs (see Table II; Bolognesi *et al*., 1997).

The high experimental resolution achieved allows the porphyrin ring substituents to be located precisely, showing that in PtrHb the heme is inserted in its specific globin crevice in an orientation that differs by a 180° rotation around the α–γ meso axis with respect to most (non-)vertebrate Hbs and Mbs (La Mar *et al*., 1978; Steigemann and Weber, 1979; Bolognesi *et al*., 1997). On the other hand, residual electron density in the CtrHb refined structure indicates that about half of the crystallized protein molecules display the conventional heme orientation. Minor modifications of the heme crevice structure and residues (Thr/GlnE11, Leu/MetF4, Met/LeuFG4 and Ala/ValH11) may account for the heme binding modes displayed by the two proteins. It has, however, been noticed that partial incorrect heme insertion can result from protein expression in *Escherichia coli* (Shen *et al*., 1993, 1997).

The rear face of the heme pocket (as shown in Figure 2) is built by the G- and H-helices, which match rather regularly the globin fold topology (Figure 2B and C). Nevertheless, the H-helix is partly bent and shows a deletion of ~10 C-terminal residues, as compared with sperm whale Mb. A bend of the H-helix in the last two turns and its local deviation from α-helical parameters underline a structural adaptation that provides closer contacts to the porphyrin ring in the inner part of the proximal side. In particular, residue ValH18 (present in both trHbs) is brought into contact with the heme next to the γ methinic bridge, at ~4 Å from the HisF8 imidazole ring (see Figure 5).

Inspection of the overall protein structure shows that an evident protein cavity/tunnel, delimited by the AB region and by helices E, G and H, is present in both trHbs. The cavity (65 $Å^3$ volume in PtrHb and 200 $Å^3$ in CtrHb, respectively, as defined by a 1.4 Å radius probe) is essentially lined by seven hydrophobic residues (LeuA15, ValB2, Thr/ValB6, LeuE15, LeuE19, LeuG12 and LeuG16). In CtrHb these residues build a continuous tunnel, ~5 Å in diameter, connecting the protein surface to an inner region very close to the heme distal site. Simple modeling shows that, besides a strong conservation of the tunnel apolar residues in the aligned trHb sequences, contained side chain fluctuations can also provide a similar solvent to heme route in PtrHb, where the cavity, as observed in the crystal, is not fully open to solvent. Such a structural feature, never observed before in (non-)vertebrate Hbs, may have substantial implications for ligand diffusion and binding processes in trHbs.

The three-dimensional stuctures of trHbs from *P.caudatum* and *C.eugametos* are in general agreement with a recent solution ^{1}H NMR investigation on the heme crevice structure of cyano-met *N.commune* trHb, which

identified the presence of PheCD1, GlnE7 and HisF8 residues together with the G- and H-helices and the FG corner (Yeh *et al*., 2000). Concerning the heme proximal region, a α-helical structure localized between the F4 and F9 residues was detected, displaying, however, an unusual increased flexibility (or significantly decreased stability) as compared with conventional (non-)vertebrate Hbs. Such an observation may be indicative of the presence of a pre-F loop also in *N.commune* Hb, which displays a sequence pattern of key residues in the EF and pre-F loop regions, including the Gly motifs, in full agreement with those of the trHbs described here (see Figure 1).

Conclusions

A structure-based alignment of trHb amino acid sequences with respect to sperm whale Mb and *Vitreoscilla* sp. Hb is reported in Figure 1 to highlight features that may be considered as a structural fingerprint of this protein homology subfamily. In the preceding sections we have underlined the structural role played in PtrHb and in CtrHb by the Gly-containing sequence motifs located next to the protein N-terminus and at both termini of the pre-F proximal loop. Inspection of Figure 1 confirms that the three Gly motifs are well conserved in the known trHb sequences, being virtually unknown in other globin structures (Bashford *et al*., 1987; Kapp *et al*., 1995). Next, the invariance of a Phe residue at the topological site E14, paired to the occurrence of Trp or Tyr residues in the pre-F loop (position 59 in Figure 1), appears to be a requirement imposed on trHbs by the lack of a heme proximal helix as well as a structural support for the wide pre-F loop. The strict Phe conservation at site E14, therefore, is reminiscent of the solvent shield role ascribed to PheCD1 in (non-)vertebrate Hbs. A further related and unexpected finding from the structure-based alignment is that some trHbs may host residues other than Phe at the CD1 site (Tyr in *M.tuberculosis* HbO, *Mycobacterium leprae* and *Corynebacterium diphtheriae* trHbs, or Leu in *B.pertussis* and *Thiobacillus ferroxidans* trHbs). Residue substitutions at site CD1 are very uncommon, being present essentially in some human Hb mutants (Antonini and Brunori, 1971; Bashford *et al*., 1987; Kapp *et al*., 1995; Bolognesi *et al*., 1997). From a functional viewpoint it may be remarked that four of the trHbs that do not bear a PheCD1 residue display Ala at the distal E7 site (Ser in *T.ferroxidans* trHb).

Comparison of amino acid sequences indicates that the extremely short Hb (109-residue) recently isolated from the nemertean worm *Cerebratulus lacteus* (Vandergon *et al*., 1998) may not fit completely within the trHb structural pattern described here, due to additional deletions at the N-terminus and residue insertions in the CD–D region, which may require further and substantial modification of the trHb fold. Nevertheless, the localization of Gly residues in the EF–F regions suggests that an extended pre-F loop may be present also in this Hb from a multicellular organism.

The *P.caudatum* globin gene has a single very short intron (29 bp) at position F3.0 (intron inserted between residue F2 and F3 codons), whereas *C.eugametos* displays introns at positions B6.0 (181 bp), E20.0 (306 bp) and F10.0 (755 bp) (Couture *et al*., 1994; Yamauchi *et al*., 1995). In contrast, the *C.lacteus* Hb gene has two introns, at positions B12.2 (intron inserted after the second base of the codon for residue B12) and G7.0; both intron positions are strongly conserved in the (non-)vertebrate globin genes (Hankeln *et al*., 1997; Dewilde *et al*., 1998). Such an observation, together with the unexpected finding of a two-over-two α-helical sandwich fold in trHbs, may support the hypothesis of an independent evolutionary origin for unicellular organism Hbs (Takagi, 1993).

As a final consideration we note that careful trimming of a conventional globin fold (horse Mb) at both termini has previously provided a heme-binding, active, 'mini-myoglobin', whose size (comprising residues B13–H16) roughly reflects the seal Mb central exon stretch (De Sanctis *et al*., 1988). The composite globin fold editing observed in the trHbs, however, indicates that the constraints imposed by a functional heme crevice on a shortened globin chain are complex and carefully distributed throughout the entire structure. How the modified fold is related to the functional properties remains to be elucidated, and will be of great significance for understanding the selective pressure that maintains trHbs in these organisms, sometimes in the presence of distinct Hb types.

Materials and methods

Crystallization of PtrHb and CtrHb

A synthetic *P.caudatum* cDNA was constructed from 15 oligonucleotides, using the codon frequency of *E.coli* according to the method of Ikehara *et al.* (1984). The gene was expressed in *E.coli* and purified as described (Dewilde *et al*., 1998). Recombinant PtrHb was crystallized by vapor diffusion techniques, at a protein concentration of 15 mg/ml. The protein droplet was equilibrated against a reservoir containing 35% ammonium sulfate and 50 mM sodium acetate pH 5.5, at 4°C. Large prismatic crystals (~0.15 × 0.15 × 0.35 mm^3) grew in ~1 week. The crystals were stored in 70% ammonium sulfate and 50 mM sodium acetate pH 5.5, and transferred to the same solution, supplemented with 20% glycerol, immediately before data collection at 100 K. The crystals grown belong to the tetragonal space group $P4_3$ and accommodate one PtrHb molecule per asymmetric unit; the crystallographic constants are reported in Table I.

CtrHb used for the crystallographic analysis was expressed as the sequence segment Ser22–Gln142, deleting the N-terminal chloroplast import sequence (lower case letters in Figure 1), and purified as described (Couture and Guertin, 1996). Crystal growth was achieved through vapor diffusion techniques; the protein solution at 40 mg/ml concentration was equilibrated against a reservoir solution containing 55% ammonium sulfate, 50 mM sodium acetate and 5 mM potassium cyanide pH 6.0, at 4°C. Rosettes of small crystals appeared after 1 week. Single tabular crystals (~0.02 × 0.05 × 0.10 mm^3) were mechanically isolated from the rosettes and used for data collections at room temperature in sealed capillaries. The heavy atom soaking conditions employed were: 15 mM $UO_2(CH_3COO)_2$ pH 6.5, 2 days; 5 mM K_2PtCl_4 pH 6.5, 1 day. For the high resolution native data collection at 100 K, CtrHb crystals were transferred to a solution with 65% ammonium sulfate, 5 mM potassium cyanide, 50 mM Tris pH 7.5, supplemented with 20% ethylene glycol as cryoprotectant. CtrHb crystallizes in the orthorhombic space group $P2_12_12_1$, with one molecule per asymmetric unit. The unit cell parameters are reported in Table I.

Structure determination and refinement

The PtrHb structure was solved by MAD techniques, based on the heme Fe atom anomalous signal, at the ESRF synchrotron source, at 100 K. Diffraction data were processed using DENZO, SCALEPACK and programs from the CCP4 suite (CCP4, 1994; Otwinoski and Minor, 1997). MAD phases were determined at 2.7 Å resolution using SOLVE (Terwilliger and Berendzen, 1999) with a figure of merit of 0.87 (Table I). The resulting electron density map clearly displayed almost all the main molecular features and residues. The program wARP (Perrakis *et al*., 1999) was used to extend and refine phases to 1.54 Å resolution and for

automated model building of main chain and side chains. The molecular model was subsequently checked manually with O (Jones *et al*., 1991) and briefly refined at the maximum resolution, using CNS (Brünger *et al*., 1998) and REFMAC (Murshudov *et al*., 1997). The final model contains 116 residues and 207 water molecules (R-factor = 13.3% and R_{free} = 18.3%, respectively; Table I), with ideal stereochemical parameters (Engh and Huber, 1991).

CtrHb structure was solved by MIR techniques, using two heavy atom derivative datasets collected in-house (Cu Kα radiation, room temperature). Diffraction data were processed using DENZO, SCALEPACK and programs from the CCP4 suite (CCP4, 1994; Otwinoski and Minor, 1997). MIR phases were determined at 3.2 Å resolution, using PHASES (Furey and Swaminathan, 1997), and were improved by SHARP (de La Fortelle and Bricogne, 1997) and by solvent flattening (SOLOMON program; Abrahams and Leslie, 1996) at 3.1 Å resolution (Table I). The initial map allowed the tracing of >50% of the model, using O (Jones *et al*., 1991). The structure was subsequently refined with CNS (Brünger *et al*., 1998), using a high resolution data set collected at ESRF, at 1.8 Å resolution, at 100 K. The final model contains 121 residues and 186 water molecules (R-factor = 17.6% and R_{free} = 21.1%, respectively; Table I), with ideal stereochemical parameters (Engh and Huber, 1991). Both trHb structures and structure factors have been deposited with the Protein Data Bank (codes 1dlw and 1dly for PtrHb and CtrHb, respectively).

Acknowledgements

We wish to thank V.Stojanof and G.Leonard for support during data collection at ESRF. We are also grateful to Professors M.Brunori, R.Huber and A.Riggs for helpful discussions. M.-L.Van Hauwaert is acknowledged for skilful technical assistance. This work was supported by grants of the Italian Ministry for University and Scientific-Technological Research (MURST Program 'Biologia Strutturale, 1997–1999'), by the CNR Target Oriented Project 'Biotecnologie', by the ASI grant ARS-98-174, by the National Science and Engineering Research Council of Canada, grant 06P0046306, and by the Fund for Scientific Research – Flanders (Belgium) (FWO) (project no. G.2023.94 and G.0314.00). S.D. is a postdoctoral fellow of the latter fund.

References

Abrahams,J.P. and Leslie,A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr. D*, **52**, 30–42.

Antonini,E. and Brunori,M. (1971) *Hemoglobin and Myoglobin in their Reactions with Ligands*. Elsevier North Holland, Amsterdam, The Netherlands, p. 70.

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.

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.

Bolognesi,M., Boffi,A., Coletta,M., Mozzarelli,A., Pesce,A., Tarricone,C. and Ascenzi,P. (1999a) Anticooperative ligand binding properties of recombinant ferric *Vitreoscilla* homodimeric hemoglobin: a thermodynamic, kinetic and X-ray crystallographic study. *J. Mol. Biol.*, **291**, 637–650.

Bolognesi,M., Rosano,C., Losso,R., Borassi,A., Rizzi,M., Wittenberg,J.B., Boffi,A. and Ascenzi,P. (1999b) Cyanide binding to *Lucina pectinata* hemoglobin I and to sperm whale myoglobin: an X-ray crystallographic study. *Biophys. J.*, **77**, 1093–1099.

Brünger,A.T. *et al*. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

Brunori,M., Cutruzzolá,F., Savino,C., Travaglini-Allocatelli,C., Vallone,B. and Gibson,Q.H. (1999a) Structural dynamics of ligand diffusion in the protein matrix: a study on a new myoglobin mutant Y(B10)Q(E7)R(E10). *Biophys. J.*, **76**, 1259–1269.

Brunori,M., Cutruzzolá,F., Savino,C., Travaglini-Allocatelli,C., Vallone,B. and Gibson,Q.H. (1999b) Does picosecond protein dynamics have survival value? *Trends Biochem. Sci.*, **24**, 253–255.

Chu,K., Vojtchovsky,J., McMahon,B., Sweet,R., Berendzen,J. and Schlichting,I. (2000) Structure of a ligand-binding intermediate in wild-type carbonmonoxy myoglobin. *Nature*, **403**, 921–923.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

Couture,M. and Guertin,M. (1996) Purification and spectroscopic characterization of a recombinant chloroplastic hemoglobin from the green unicellular alga *Chlamydomonas eugametos*. *Eur. J. Biochem.*, **242**, 779–787.

Couture,M., Chamberland,H., St-Pierre,B., Lafontaine,J. and Guertin,M. (1994) Nuclear genes encoding chloroplast hemoglobins in the unicellular green alga *Chlamydomonas eugametos. Mol. Gen. Genet.*, **243**, 185–197.

Couture,M., Das,T.K., Lee,H.C., Peisach,J., Rousseau,D.L., Wittenberg, B.A., Wittenberg,J.B. and Guertin,M. (1999a) *Chlamydomonas* chloroplast ferrous hemoglobin. Heme pocket structure and reactions with ligands. *J. Biol. Chem.*, **274**, 6898–6910.

Couture,M., Yeh,S., Wittenberg,B.A., Wittenberg,J.B., Ouellet,Y., Rousseau,D.L. and Guertin,M. (1999b) A cooperative oxygen-binding hemoglobin from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **96**, 11223–11228.

Das,T.K., Couture,M., Lee,H.C., Peisach,J., Rousseau,D.L., Wittenberg, B.A., Wittenberg,J.B. and Guertin,M. (1999) Identification of the ligands to the ferric heme of *Chlamydomonas* chloroplast hemoglobin: evidence for ligation of tyrosine-63 (B10) to the heme. *Biochemistry*, **38**, 15360–15368.

de La Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.

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.

Dewilde,S. *et al*. (1998) Structural, functional and genetic characterization of *Gastrophilus* hemoglobin. *J. Biol. Chem.*, **273**, 32467–32474.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A*, **47**, 392–400.

Ermler,U., Siddiqui,R.A., Cramm,R. and Friedrich,B. (1995) Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 Å resolution. *EMBO J.*, **14**, 6067–6077.

Furey,W. and Swaminathan,S. (1997) PHASES-95: a program package for processing and analysing diffraction data from macromolecules. *Methods Enzymol.*, **277**, 590–620.

Hankeln,T., Friedl,H., Ebersberger,I., Martin,J. and Schmidt,E.R. (1997) A variable intron distribution in globin genes of *Chironomus*: evidence for recent intron gain. *Gene*, **205**, 151–160.

Hardison,R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.*, **201**, 1099–1117.

Hempstead,P.D., Yewdall,S.J., Fernie,A.R., Lawson,D.M., Artymiuk, P.J., Rice,D.W., Ford,G.C. and Harrison,P.M. (1997) Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *J. Mol. Biol.*, **268**, 424–448.

Holm,L. and Sander,C. (1993) Structural alignment of globins, phycocyanins and colicin A. *FEBS Lett.*, **315**, 301–306.

Ikehara,M. *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.

Imai,K. (1999) The haemoglobin enzyme. *Nature*, **401**, 437.

Iwaasa,H., Takagi,T. and Shikama,K. (1989) Protozoan myoglobin from *Paramecium caudatum*. Its unusual amino acid sequence. *J. Mol. Biol.*, **208**, 355–358.

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. A*, **47**, 110–119.

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.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

La Mar,G.N., Overkamp,M., Sick,H. and Gersonde,K. (1978) Proton nuclear magnetic resonance hyperfine shifts as indicators of tertiary structural changes accompanying the Bohr effect in monomeric insect hemoglobins. *Biochemistry*, **17**, 352–361.

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.

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.

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.

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 relationships. *Mol. Biol. Evol.*, **13**, 324–333.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.

Otwinoski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Perrakis,A., Morris,R. and Lamzin,V. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.*, **6**, 458–463.

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.

Potts,M., Angeloni,S.V., Ebel,R.E. and Bassam,D. (1992) Myoglobin in a cyanobacterium. *Science*, **256**, 1690–1691.

Shen,T.J., Ho,N.T., Simplaceanu,V., Zou,M., Green,B.N., Tam,M.F. and Ho,C. (1993) Production of unmodified human adult hemoglobin in *Escherichia coli. Proc. Natl Acad. Sci. USA*, **90**, 8108–8112.

Shen,T.J., Ho,N.T., Zou,M., Sun,D.P., Cottam,P.F., Simplaceanu,V., Tam,M.F., Bell,D.A.,Jr and Ho,C. (1997) Production of human normal adult and fetal hemoglobins in *Escherichia coli. Protein Eng.*, **10**, 1085–1097.

Steigemann,W. and Weber,E. (1979) Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution. *J. Mol. Biol.*, **127**, 309–338.

Takagi,T. (1993) Hemoglobins from single-celled organisms. *Curr. Opin. Struct. Biol.*, **3**, 413–418.

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.

Terwilliger,T.C. and Berendzen,J. (1999) Automated structure solution for MIR and MAD. *Acta Crystallogr. D*, **55**, 849–861.

Thorsteinsson,M.V., Bevan,D.R., Potts,M., Dou,Y., Eich,R.F., Hargrove,M.S., Gibson,Q.H. and Olson,J.S. (1999) A cyanobacterial hemoglobin with unusual ligand binding kinetics and stability properties. *Biochemistry*, **38**, 2117–2126.

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.

Yamauchi,K., Tada,H. and Usuki,I. (1995) Structure and evolution of *Paramecium* hemoglobin genes. *Biochim. Biophys. Acta*, **1264**, 53–62.

Yang,J., Kloek,A.P., Goldberg,D.E. and Matthews,F.S. (1995) The structure of *Ascaris* hemoglobin domain I at 2.2 Å resolution: molecular features of oxygen avidity. *Proc. Natl Acad. Sci. USA*, **92**, 4224–4228.

Yeh,D.C., Thorsteinsson,M.V., Bevan,D.R., Potts,M. and La Mar,G.N. (2000) Solution ^{1}H NMR study of the heme cavity and folding topology of the abbreviated chain 118-residue globin from the cyanobacterium *Nostoc commune*. *Biochemistry*, **39**, 1389–1399.

Received March 21, 2000; accepted April 4, 2000