Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution structure of human cytoglobin: identification of extra N- and C-termini and a new dimerization mode

Masatomo Makino,[a,b] Hiroshi Sugimoto,[a]* Hitomi Sawai,[a,b] Norifumi Kawada,[c] Katsutoshi Yoshizato[d] and Yoshitsugu Shiro[a]

[a]Biometal Science Laboratory, Harima Institute, RIKEN SPring-8 Center, Harima Institute, Japan, [b]Department of Life Science, Graduate School of Life Science, University of Hyogo, Japan, [c]Department of Hepatology, Graduate School of Medicine, Osaka City University, Japan, and [d]Department of Biological Science, Graduate School of Science, Hiroshima University, Japan

Correspondence e-mail: sugimoto@spring8.or.jp

Received 26 January 2006
Accepted 17 April 2006

PDB Reference: human cytoglobin, 2dc3, r2dc3sf.

Cytoglobin (Cgb) is a recently discovered member of the vertebrate haem-containing globin family. The structure of a new crystal form of wild-type human Cgb (space group *C*2) was determined at a resolution of 1.68 Å. The results show the presence of an additional helix in the N-terminal residues (4–20) prior to the *A* helix and an ordered loop structure in the C-terminal region (168–188), while these extended peptides were invisible owing to disorder in the previously reported structures using a $P3_221$ crystal at a resolution of 2.4 Å. A detailed comparison of the two crystal structures shows differences in the conformation of the residues (*i.e.* Arg84) in the haem environment owing to a different dimeric arrangement.

1. Introduction

Cytoglobin (Cgb), a member of the vertebrate haem-containing globin family, was first discovered using a proteome approach for the molecular analysis of rat stellate cells induced by the administration of thioacetamide (Kawada *et al.*, 2001). A recent study revealed that its mRNA is expressed in a wide range of organs in both human and mouse tissues (Burmester *et al.*, 2002) and that its expression is up-regulated under conditions of hypoxia and certain types of stress (Schmidt *et al.*, 2004). A comparison of its primary structure with those of myoglobin (Mb) and haemoglobin (Hb) indicate that Cgb shares only 25% identity with Mb and Hb, but that some amino-acid residues are highly conserved (Pesce *et al.*, 2002). In the case of Mb and Hb, the conserved residues are located around the haem active sites and are responsible for the binding of external ligands to the haem iron. It has therefore been suggested that Cgb is possibly involved in O_2 binding, although its precise physiological role has not yet been established. In addition, Cgb is composed of 190 amino-acid residues, making it longer than other globins, which typically contain 140–160 residues (Trent & Hargrove, 2002). In the sequence alignment, the extra peptides appear to be present in the N- and C-terminal regions.

Recently, the structures of the wild type (Sugimoto *et al.*, 2004) and a mutant (C38S/C83S; de Sanctis *et al.*, 2004) of recombinant human Cgb have been crystallographically determined for the ferric form at resolutions of 2.4 and 2.1 Å, respectively. The overall backbone structures in these crystals exhibit the classical three-over-three α-helical sandwich globin fold that is observed in other globin proteins, including not only vertebrate Mb and Hb but also Hbs from a variety of

other sources. However, the structure of the active site of ferric Cgb was significantly different from those of traditional Hb and Mb. In ferric Cgb, the imidazole group of His81(*E*7) directly coordinates to the haem iron as a sixth axial ligand to form a low-spin complex, while a water molecule is present at this position in vertebrate Mb and Hb, forming a ferric high-spin complex. The ferrous form of Cgb also has a hexa-coordinated haem iron with two histidyl imidazole groups at axial positions, as has been shown by several spectroscopic studies (Trent & Hargrove, 2002; Sawai *et al.*, 2003). Some external ligands such as O_2, CO and NO are able to bind to the haem iron after dissociation of the endogenous His81(*E*7) imidazole ligand from the sixth position. Such structural differences between Cgb and Mb are reflected in some of the kinetic properties in ligand binding (Trent & Hargrove, 2002), the auto-oxidation rate of the oxygen (O_2) complex and its redox properties (Sawai *et al.*, 2003).

From the viewpoint of protein crystallography, a detailed structural analysis of globin proteins using high-resolution data is of interest for comparing their divergent oxygen-binding properties and to define the electron density of the N- and C-terminal regions, which have not been observed in previous studies. In the present study, a search for higher quality crystals resulted in the isolation of a novel crystal form of Cgb with significantly improved diffraction to a resolution of 1.68 Å. The characteristic properties of Cgb are discussed based on the high-resolution structure of Cgb.

2. Experimental procedures

2.1. Purification and crystallization

Recombinant human Cgb with a 6×His tag was expressed in *Escherichia coli* overproducing strain BL21(DE3) as described previously (Sawai *et al.*, 2003; Sugimoto *et al.*, 2004). The human Cgb was purified by Ni–NTA (Qiagen) affinity chromatography and Mono Q (Amersham Pharmacia) anion-exchange chromatography. The N-terminal His tag was cleaved by treatment with biotinylated thrombin (Novagen) and was captured using streptavidin agarose (Novagen) after cleavage. The purified protein was oxidized to the ferric form by the addition of an excess of potassium ferricyanide and then desalted on a HiPrep 26/10 desalting column (Amersham Pharmacia) previously equilibrated with 20 m*M* Na HEPES pH 8.0. The sample was concentrated by centrifugation in YM-10 (Millipore) to a concentration of between 30 and 40 mg ml^{-1}. The purified Cgb contains three amino acids (Gly, Ser, His) from a cloning artifact in the N-terminus.

2.2. Crystallization and data collection

Crystallization of human Cgb was carried out by the sitting-drop vapour-diffusion method at 293 K. In all experiments, the crystallization drops, containing 1 µl protein solution and 1 µl precipitant solution, were equilibrated against 70 µl reservoir solution (the same as the precipitant solution). Crystals of human Cgb were obtained using 1.4 *M* sodium acetate, 100 m*M* sodium cacodylate pH 6.5, 5 m*M* KCN. The crystals were soaked in the reservoir solution, which contained 25% xylitol as a cryoprotectant. X-ray diffraction data were collected to a resolution of 1.68 Å at beamline BL45XU at SPring-8 using a Jupiter 210 (Rigaku) detector (Yamamoto *et al.*, 1998). The crystal belongs to space group *C*2, with unit-cell parameters $a = 78.5$, $b = 52.7$, $c = 83.4$ Å, and includes two Cgb monomers per asymmetric unit, with a solvent content of 39.7%. All diffraction data were integrated and scaled using *HKL*2000 (Otwinowski & Minor, 1997). Data-collection and processing statistics are summarized in Table 1.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
X-ray source (beamline)	SPring-8 (BL45PX)
Wavelength (Å)	1.000
Resolution (Å)	50–1.68 (1.74–1.68)
Space group	*C*2
Unit-cell parameters (Å, °)	$a = 78.5$, $b = 52.8$, $c = 83.4$, $\beta = 89.8$
Observed reflections	75822
Unique reflections	39043
R_{sym}† (%)	5.6 (28.6)
$I/\sigma(I)$	20.9 (4.0)
Redundancy	4.9 (4.1)
Completeness (%)	99.9 (99.5)
Refinement statistics and model quality	
Resolution (Å)	20–1.68
Total atoms	3142
R_{work}/R_{free}‡ (%)	14.0/18.1
R.m.s.d. bonds (Å)	0.018
R.m.s.d. angles (°)	1.6
Ramachandran plot§	
Most favoured regions (%)	94.4
Additional allowed regions (%)	5.6
Average *B* factors (Å^2)	
Protein	15.6
Main chain	14.0
Side chain	17.2
Haem	9.5
Water	26.9

† $R_{sym} = \sum_{hkl} |I(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} |I(hkl)|$, where $\langle I(hkl)\rangle$ is the average of $I(hkl)$. ‡ R factor $= \sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)| / \sum_{hkl} F_{obs}(hkl)$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. R_{free} is the R factor calculated using 5% of the randomly selected reflections not included for refinement. § The Ramachandran plot was constructed using *PROCHECK* (Laskowski *et al.*, 1993).

2.3. Structure-determination and refinement procedures

The initial phase of the *C*2 form was determined by molecular replacement using the atomic coordinates for residues 21–171 of the previously reported Cgb in space group $P3_221$ (PDB code 1v5h) as a search molecule using the *MOLREP* program (Vagin & Teplyakov, 1997). The solution from *MOLREP* resulted in an R value of 44.8% and a correlation coefficient of 0.523 in the resolution range 8.0–3.0 Å. The orientations and positions of two molecules were found in an asymmetric unit and were improved by rigid-body refinement.

5% of the randomly selected reflections were designated as test reflections for use in the R_{free} cross-validation method (Brünger, 1992) and were used throughout the refinement. In the first stage of the model refinement using *CNS* (Brünger *et al.*, 1998), a simulated-annealing protocol was employed. The

model was further refined with multiple rounds of positional and individual *B*-factor refinement in the resolution range 20–1.68 Å followed by manual model rebuilding using *O* (Jones *et al.*, 1991). Non-crystallographic symmetry restraints were not used in the refinement process. The C-terminal regions of subunit *A*, which were not determined in the $P3_221$ crystal, were built manually into calculated $2F_o - F_c$ maps contoured at 1.1σ and $F_o - F_c$ maps contoured at 2.4σ. After several refinement cycles, anisotropic refinement was applied using *REFMAC*5 (Murshudov *et al.*, 1997). At this stage, the $F_o - F_c$ map showed positive density for the N-terminal region of subunit *B*, which enabled 4–20 residues to be added to the model. The final model contained residues 17–188 of subunit *A*, resiues 4–171 of subunit *B* and 265 water molecules. The residues Met30, Ser103, Ser104 and Ser141 of subunit *A* and Gln62, Met72, Ser128 and Ser160 of subunit *B* show an alternative conformation. The figures were generated using *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Structure determination

Recombinant human Cgb was crystallized using sodium acetate as a precipitant in conjunction with the sitting-drop vapour-diffusion method. The crystal belongs to space group *C*2, contains two monomers per asymmetric unit and diffracts to a resolution of 1.68 Å. The structure of the new crystal form (*C*2-Cgb) was determined by molecular replacement and refined with anisotropic temperature factors and without the use of non-crystallographic symmetry restraints. The resulting electron-density map enabled the modelling of the N- and C-terminal regions. The *C*2-Cgb molecular packing is different from those for previously reported crystals that belong to space group $P3_221$ for the wild type or $P2_12_12_1$ for the C38S/C83S mutant.

3.2. Overall structure

The overall structure of *C*2-Cgb is shown in Fig. 1(*a*). The core portion of the monomer structure is a classical globin fold, the same as that for the previously reported $P3_221$ crystal ($P3_221$-Cgb; Fig. 1*b*) and other globins. In a quantitative comparison between the *C*2-Cgb and $P3_221$-Cgb structures, the root-mean-square difference is 0.68 Å for the 151 C^α atoms of residues 21–171. With respect to the backbone structure, small but significant differences were observed in the *AB* and the *GH* corners. Since the *AB* corner and helix *E* are involved in dimer formation of Cgb (see below), structural differences observed in this region can be attributed to the dimer orientation between *C*2-Cgb and $P3_221$-Cgb.

As the case for $P3_221$-Cgb, *C*2-Cgb also forms a homodimer through two intermolecular disulfide bonds between Cys38 at the *AB* corner and Cys83 at the centre of helix *E*. However, in the superposition of monomers from different crystal forms, the orientation of the dimer partner differs by about 90°. The total buried surface area as a result of dimerization in *C*2-Cgb is 643 $Å^2$ per subunit, an approximately twofold increase in area compared with $P3_221$-Cgb, and barely passes the lower end of what is expected for a stable dimer (Janin *et al.*, 1988). A remarkable diversity in quaternary structure is observed between invertebrate and primitive vertebrate Hbs despite the conserved fold of the subunits (Royer *et al.*, 2001). It is interesting that the resultant dimer (quaternary) structure of *C*2-Cgb is very similar to the dimer of sea lamprey Hb from *Petromyzon marinus* (Fig. 1*c*; Heaslet & Royer, 1999). Lamprey Hb shares 26% sequence identity to human Cgb, which is the highest identity among vertebrate globin protein members. A previous report of the structure analysis of $P3_221$-Cgb has pointed out that the backbone conformation of the *E* and *F* helices and dimer interface of Cgb bears the best resemblance to those of lamprey Hb, except for the quaternary structure. The similarity in primary to quaternary structures revealed by the present analysis allows us to propose a similar role of dimerization in Cgb and lamprey Hb in exerting their function. For example, lamprey Hb is in an equilibrium state between a low-affinity oligomer and a high-affinity monomer, which contributes to its cooperativity. A detailed structural comparison between the deoxy (Heaslet & Royer, 1999) and cyanide derivatives (Hendrickson *et al.*, 1988) of lamprey Hb revealed that oligomerization causes a movement of the first half of the *E* helix. It has been proposed that the O_2 affinity of lamprey Hb can be regulated through movement in the haem distal side. The dodecameric Hbs (also known as the giant Hb) found in annelids exhibit strong cooperative oxygen binding and use an *E* helix for their heterodimer interface (Strand *et al.*, 2004). On the other hand, it was recently reported that Cgb also exhibits cooperativity in O_2 binding (Fago *et al.*, 2004), in which the O_2 ligation at one haem affects the affinity of another haem group. On the basis of the discussion relative to lamprey Hb and other cooperative Hbs, the cooperative behaviour in Cgb might be accomplished through the dimerization interface, *i.e.* the *AB* corner and the *E* helix. Although the reducing environment *in vivo* prevents disulfide-bond formation, the previous and present dimeric structures of Cgb might reflect the interactions in living cells even if disulfide bonds are not formed. Further studies will be necessary to evaluate the relationship between quaternary structure and cooperative binding for this molecule.

3.3. N- and C-terminal structures of *C*2-Cgb

Cgb contains additional sequences in the N- and C-terminal regions. The C-terminal region of subunit *A* of the dimer and the N-terminal region of subunit *B* show interpretable electron density, despite their very similar structure in the globin core region. These regions and the *GH* corner of subunit *B* exhibit a relatively high temperature factor. Electron density is missing in the N-terminal region of subunit *A* and in the C-terminal region of subunit *B*. The final refined structure contains 172 residues (17–188) for subunit *A* and 168 residues (4–171) for subunit *B*, with 265 water molecules and one acetate ion in the asymmetric unit and one haem group per subunit.

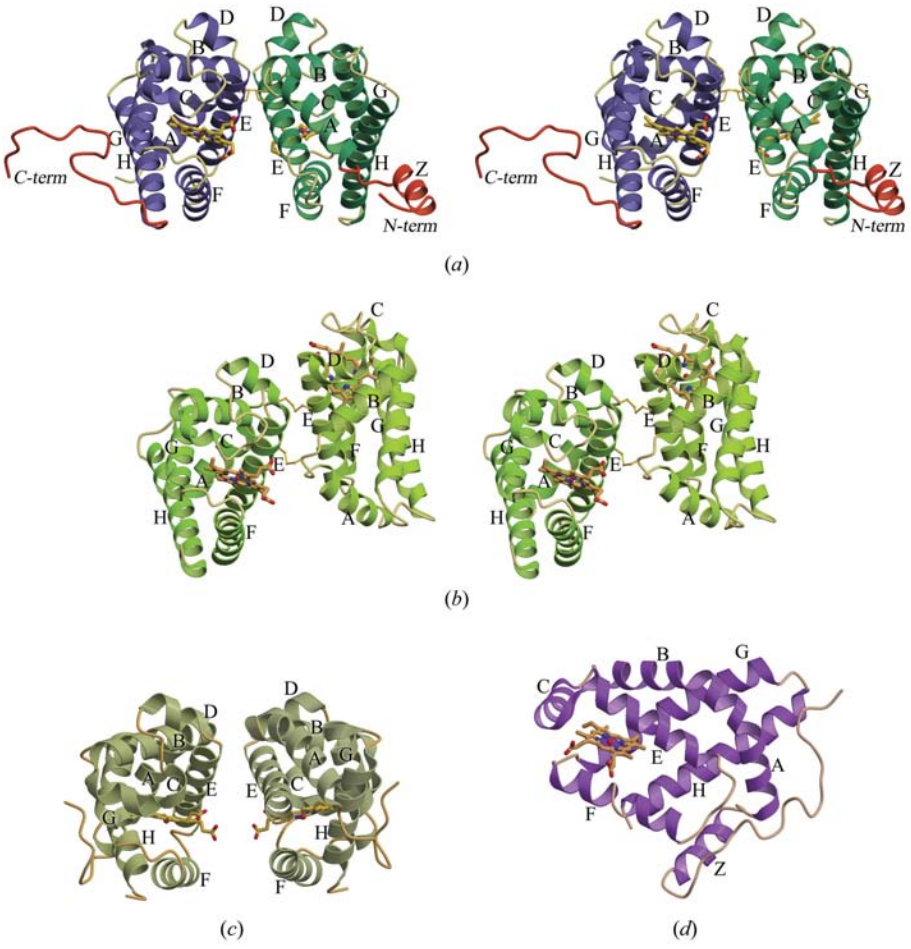

Figure 1
(*a*) Stereo diagram of dimeric structure of Cgb. The core portions of subunit *A* and subunit *B* of *C*2-Cgb are coloured blue and green, respectively. N- and C-terminal regions are coloured red. The α-helices are labelled alphabetically according to the nomenclature for the classic globin fold. The haem group and Cys residues located at the *B*2 and *E*9 positions are represented by a ball-and-stick model. The two disulfide bonds in Cgb are coloured yellow. For comparison, (*b*) the $P3_221$-Cgb dimer (PDB code 1v5h), whose left subunit is represented in the same orientation of subunit *A* of *C*2-Cgb, (*c*) the lamprey Hb dimer (PDB code 3lhb) and (*d*) the sensor domain of the haem-based aerotaxis transducer (HemAT) from *B. subtilis* (PDB code 1or4) are shown.

Owing to these interactions of the Cgb core part with the *Z* helix or the C-terminal coil, structural differences between subunits *A* and *B* were evident in the *CD* and *GH* regions. When two subunits are superimposed for the 151 C^{α} atoms with r.m.s.d. 0.50 Å, the deviation for the *CD* and *GH* corners are 1.12 and 1.30 Å, respectively. The *GH* corner of subunit *B* exhibits a high degree of conformational flexibility, while the corresponding region for subunit *A* is well ordered because of interactions between the *GH* corner of subunit *B* and the C-terminal coil of the other adjacent dimer in the crystal. In other words, various types of interactions between the C-terminal loop of subunit *A* and the *GH* corner mutually stabilize their structure in the crystal.

To examine the reason why the *Z* helix is not observed in subunit *A*, subunit *B* was superimposed on subunit *A*. The putative *Z* helix in subunit *A* would collide with the *F*-helix region (residues 90–102) of subunit *A*′ in the crystal packing. The postulated C-terminal random coil for subunit *B* obtained by the same operation as for the *Z* helix is disturbed in the *A* helix of subunit *B*′. This is likely to be a consequence of the fact that the N-terminal region of subunit *A* and the C-terminal region of subunit of *B* is deformed or disordered in the crystal.

We applied the same modelling procedure to $P3_221$-Cgb and found that the C-terminal loop penetrates into the *B* helix of the symmetry-related molecule. This is the reason why the C-terminus was invisible in $P3_221$-Cgb. On the other hand, the *Z* helix would not be involved in any collisions but would be completely free from any interactions with other symmetry-related molecules in $P3_221$ crystals. It is possible that human Cgb is composed of nine α-helices (*Z*, *A*, *B*, *C*, *D*, *E*, *F*, *G* and *H*) and a C-terminal random coil, even in the solution state. Such a structure having an external helix in the pre-*A* segment is the first example of such a structure among globin proteins. One exception is the sensor domain of the haem-based aerotaxis transducer (HemAT) from *Bacillus subtilis* (Hou *et al.*, 2000, 2001), which exhibits a globin fold with the pre-*A* helix (Zhang & Phillips, 2003; Fig. 1*d*).

The C-terminal region of subunit *A*, consisting of amino-acid residues 172–188, does not form a specific secondary structure such as an α-helix or a β-strand, but a random coil structure with one type I β-turn in residues Ala177–Thr180 and a *cis*-peptide bond between Val175 and Pro176. The random coil lies in a crevice formed by subunit *A* and symmetry-related neighbouring subunits in the crystal lattice (subunits *A*′ and *B*′) and interacts with the *GH* corner of subunit *A*′ (see Fig. 2*a*).

The N-terminal pre-*A* helix region of *C*2-Cgb constitutes a helical structure consisting of nine residues. We call this new helix the '*Z* helix'. The *Z* helix of subunit *B* is in contact with the *CD* corner and the *F* helix of a symmetry-related subunit (subunit *B*′) through hydrogen bonding as well as van der Waals interactions, as was illustrated in Fig. 2(*b*), but does not make contact with its own core part.

3.4. Structure of haem pocket

When *C*2 crystals were prepared, the crystallization drop was set up in the presence of cyanide ions (CN^-). However,

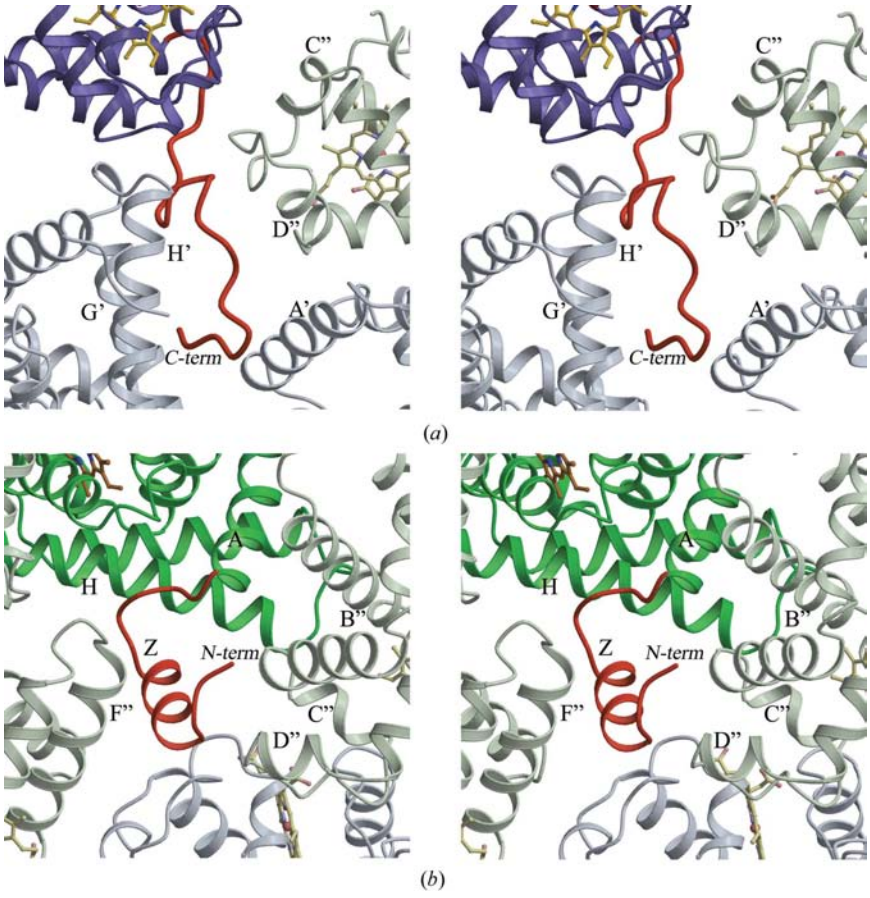

Figure 2
Stereo diagram of the crystal packing of *C*2-Cgb. The packing of Cgb molecules is shown as a ribbon model and the haem group of each molecule is drawn as a ball-and-stick model. For clarity, the α-helices for symmetry-related subunit *A* and subunit *B* are labelled with primes and double primes, respectively, and coloured grey. (*a*) The C-terminal random coil of subunit *A* makes a small intermolecular interaction with the *GH* corner of a symmetry-related molecule. (*b*) The *Z* helix of subunit *B* contacts the *CD* region and *F* helix of symmetry-related molecules.

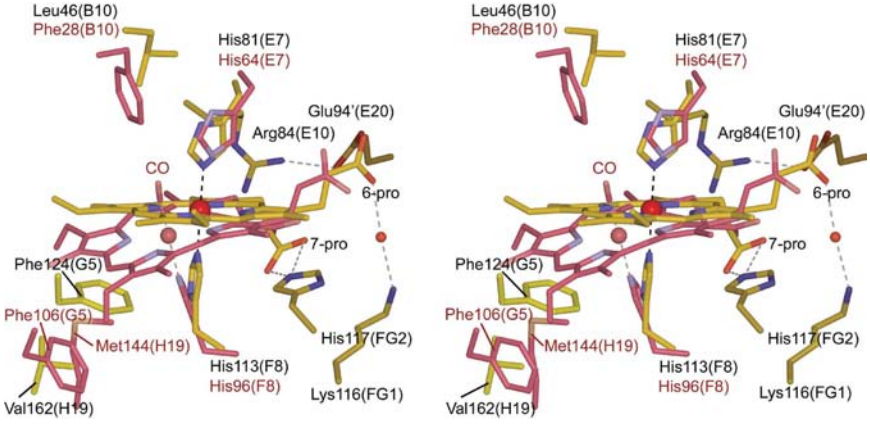

Figure 3
Superposition of *C*2-Cgb and the Ngb–CO complex (PDB code 1w92) showing the haem environment in stereo. The stick models of *C*2-Cgb and the Ngb–CO complex are coloured yellow and pink, respectively. The labels on the residues of *C*2-Cgb and Ngb complex are coloured black and red, respectively. The label in parentheses represents the helix name and the position in the helix. Arg84(*E*10) of Cgb interacts with Glu94′(*E*20) of the dimer partner. The structural change of Ngb on CO binding involves rotation of the side chain of Phe106(*G*5) and a slide of the haem. In the case of Cgb, however, steric collisions between Phe124(*G*5) and Val162(*H*19) do not allow such movements.

the sixth position of the haem iron in *C*2-Cgb is occupied by an endogenous His81(*E*7) imidazole with a Fe—$N^{\varepsilon 2}$ bond length of 2.1 Å, as is the case for the $P3_221$-Cgb. In the present case, cyanide ions might be vapourized as hydrogen cyanide during three months of crystal growth in an acidic pH. It may also be possible that the cyanide ion that was bound as a sixth ligand was removed from haem coordination during data collection because of photo-reduction of iron by the X-rays. With respect to the haem environment, several differences between *C*2-Cgb and $P3_221$-Cgb were observed in the location and position of the amino-acid residues. The butyl group of Leu46(*B*10) is rotated by 90° around the C^{β}—C^{γ} bond. In $P3_221$-Cgb, the 7-propionic acid on the pyrrole ring *A* interacts with both Lys116(*FG*1) N^{ζ} and His117(*FG*2) $N^{\varepsilon 2}$. In contrast, the 7-propionic acid of *C*2-Cgb only interacts with His117(*FG*2) $N^{\varepsilon 2}$, while Lys116(*FG*1) is in contact with the 6-propionic acid on pyrrole ring *D* *via* a water-mediated hydrogen bond (Fig. 3).

A striking difference was observed in the case of residue Arg84(*E*10), which is located in close proximity to the distal His81(*E*7) on the *E* helix. In $P3_221$-Cgb, Arg84(*E*10) exhibited significant conformational flexibility in electron density and was refined to two alternate conformations. In the major conformation with a 60% occupancy, the guanidium side chain swings outward from the haem pocket and is exposed to the solvent, while the side chain in the minor conformation, with a 40% occupancy, is in close proximity to the sixth ligand-binding position. Both conformations of Arg84(*E*10) in $P3_221$-Cgb were free from any interactions which could lead to alternate conformations. Spectroscopic analysis suggests that this residue might be involved in modulation of the ligand binding to the haem iron of Cgb (Sawai *et al.*, 2003). The conformation of Arg84(*E*10) in *C*2-Cgb is stabilized by interaction of the $N^{\eta 1}$ atom with Glu94(*E*20) $O^{\varepsilon 1}$ of the dimer partner, with a distance of 3.0 Å (Fig. 3). This intermolecular salt bridge is accomplished by a dimer orientation that is specific to *C*2-Cgb (see above).

Crystal structures obtained from the $P3_221$ and *C*2 crystal forms reveal that Cgb has two dimerization modes. Although the C38S/C83S mutant also forms a dimer in the asymmetric unit of the crystal, the disrup-

tion of the disulfide bond by the mutation of C38S/C83S results in subunit orientations that are completely different from those of $P3_221$-Cgb and $C2$-Cgb. In the $C2$ crystal form, residues located in the dimeric interface of one monomer have a direct effect on the conformations of the distal residues of the other monomer. This suggests that the equilibrium between the monomeric and dimeric forms or changes in subunit orientation might influence the ligand-association/dissociation properties of Cgb through residues on the distal side, which is part of the dimeric interface.

Vallone and coworkers recently reported the crystal structure of the CO-bound form of ferrous murine neuroglobin (Ngb; Vallone *et al.*, 2004), which is also a recently discovered member of the vertebrate globin family (Burmester *et al.*, 2000) with a hexacoordinated haem iron (Trent *et al.*, 2001). This structure might be a model for exogenous ligand binding for hexacoordinated globins, since Ngb, like Cgb, has the same fold as other globin proteins and contains a haem hexacoordinated with two His residues at the axial position of the iron in both the ferric and ferrous states. In the crystal structure of Ngb, the CO ligand occupies the sixth position after expelling the His(*E*7) imidazole, but in order to create a vacant sixth site for CO binding the haem group slides deeper into the crevice, as illustrated in Fig. 3. The repositioning of the haem causes neither a swing motion of the His(*E*7) imidazole nor conformational changes in the distal side, but induces a drastic repositioning of the Phe106(*G*5) side chain, which is located on the haem-proximal side of Ngb. Therefore, Phe(*G*5) appears to be a key residue in controlling the haem-sliding motion for exogenous ligands that bind to hexacoordinated globins. Phe(*G*5) is also present in human Cgb. To examine the structural rearrangement on ligand binding in Cgb, a modelling study was carried outbased on the present structure: *i.e.* the side chain of the Phe124(*G*5) group was shifted toward the *H* helix in the same manner as CO binding induced a structural change in Phe106(*G*5) of Ngb. However, the steric collision by juxtaposed Val162(*H*19) does not allow such a rotation in Cgb and little room is available for haem sliding in Cgb. This suggests that the mechanism of the ligand-induced conformational changes differs from that in Ngb. It is more plausible that lamprey Hb and Cgb share a common mechanism in which the structural rearrangement at the distal side is induced upon coordination of an external ligand to the haem iron.

In summary, the structure of human Cgb presented here permits a significant improvement in the accuracy of the model compared with previously reported structures and allows a better understanding of its unique features. Firstly, the overall backbone structure of Cgb exhibits a traditional globin fold with an additional helix in the pre-*A* helix region and an ordered random coil structure in the C-terminal region. This is the first example of an external helix at the pre-*A* segment in globin proteins. Secondly, the new crystal form shows an alternate dimeric arrangement that is similar to the lamprey Hb dimer. Although the physiological relevance of these two types of dimeric structures of Cgb is not clear, it is possible that a structural change in the dimer interface affects ligand-binding properties, as has been observed for lamprey Hb.

This work was supported in part by the Structural Biology Research Program at RIKEN (to YS), by the Molecular Ensemble Research Program at RIKEN (to YS) and by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Culture and Sports of Japan (to HS and YS). MM is supported by the JSPS Research Fellowship for Young Scientists.

References

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
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., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burmester, T., Ebner, B., Weich, B. & Hankeln, T. (2002). *Mol. Biol. Evol.* **19**, 416–421.
Burmester, T., Weich, B., Reinhardt, S. & Hankeln, T. (2000). *Nature (London)*, **407**, 520–523.
Fago, A., Hundahl, C., Dewilde, S., Gilany, K., Moens, L. & Weber, R. E. (2004). *J. Biol. Chem.* **279**, 44417–44426.
Heaslet, H. A. & Royer, W. E. Jr (1999). *Structure*, **7**, 517–526.
Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. & Merritt, E. A. (1988). *Proteins*, **4**, 77–88.
Hou, S., Freitas, T., Larsen, R. W., Piatibratov, M., Sivozhelezov, V., Yamamoto, A., Meleshkevitch, E. A., Zimmer, M., Ordal, G. W. & Alam, M. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 9353–9358.
Hou, S., Larsen, R. W., Boudko, D., Riley, C. W., Karatan, E., Zimmer, M., Ordal, G. W. & Alam, M. (2000). *Nature (London)*, **403**, 540–544.
Janin, J., Miller, S. & Chothia, C. (1988). *J. Mol. Biol.* **204**, 155–164.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kawada, N., Kristensen, D. B., Asahina, K., Nakatani, K., Minamiyama, Y., Seki, S. & Yoshizato, K. (2001). *J. Biol. Chem.* **276**, 25318–25323.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pesce, A., Bolognesi, M., Bocedi, A., Ascenzi, P., Dewilde, S., Moens, L., Hankeln, T. & Burmester, T. (2002). *EMBO Rep.* **3**, 1146–1151.
Royer, W. E. Jr, Knapp, J. E., Strand, K. & Heaslet, H. A. (2001). *Trends Biol. Sci.* **26**, 297–304.
Sanctis, D. de, Dewilde, S., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. & Bolognesi, M. (2004). *J. Mol. Biol.* **336**, 917–927.
Sawai, H., Kawada, N., Yoshizato, K., Nakajima, H., Aono, S. & Shiro, Y. (2003). *Biochemistry*, **42**, 5133–5142.
Schmidt, M., Gerlach, F., Avivi, A., Laufs, T., Wystub, S., Simpson, J. C., Nevo, E., Saaler-Reinhardt, S., Reuss, S., Hankeln, T. & Burmester, T. (2004). *J. Biol. Chem.* **279**, 8063–8069.
Strand, K., Knapp, J. E., Bhyravbhatla, B. & Royer, W. E. Jr (2004). *J. Mol. Biol.* **344**, 119–134.
Sugimoto, H., Makino, M., Sawai, H., Kawada, N., Yoshizato, K. & Shiro, Y. (2004). *J. Mol. Biol.* **339**, 873–885.
Trent, J. T. III & Hargrove, M. S. (2002). *J. Biol. Chem.* **277**, 19538–19545.
Trent, J. T. III, Watts, R. A. & Hargrove, M. S. (2001). *J. Biol. Chem.* **276**, 30106–30110.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vallone, B., Nienhaus, K., Matthes, A., Brunori, M. & Nienhaus, G. U. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 17351–17356.
Yamamoto, M., Kumasaka, T., Fujisawa, T. & Ueki, T. (1998). *J. Synchrotron Rad.* **5**, 222–225.
Zhang, W. & Phillips, G. N. Jr (2003). *Structure*, **11**, 1097–1110.