Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

New insight into the haemoglobin superfamily: preliminary crystallographic characterization of human cytoglobin

Daniele de Sanctis,[a] Sylvia Dewilde,[b] Alessandra Pesce,[a] Paolo Ascenzi,[c] Thorsten Burmester,[d] Thomas Hankeln,[e] Luc Moens[b] and Martino Bolognesi[a]*

[a]Department of Physics – INFM and Center for Excellence in Biomedical Research, University of Genova, Via Dodecaneso 33, I-16146 Genova, Italy, [b]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium, [c]Department of Biology, University 'Roma Tre', Viale Guglielmo Marconi 446, I-00146 Roma, Italy, [d]Institute of Zoology, Johannes Gutenberg University of Mainz, Müllerweg 6, D-55099 Mainz, Germany, and [e]Institute of Molecular Genetics, Johannes Gutenberg University of Mainz, Becherweg 32, D-55099 Mainz, Germany

Correspondence e-mail: bolognes@fisica.unige.it

Received 14 February 2003
Accepted 6 May 2003

Human cytoglobin, present in almost all tissue types, is a newly identified member of the Hb superfamily. A double mutant, having both cysteines replaced by serines, has been overexpressed in *Escherichia coli*, purified and crystallized. A highly redundant SAD data set has been collected at the haem Fe-atom absorption edge (λ = 1.720 Å) to 2.60 Å resolution. The crystals belong to the orthorhombic $P2_12_12_1$ space group, with unit-cell parameters a = 46.8, b = 73.1, c = 98.9 Å and two molecules per asymmetric unit. The anomalous difference Patterson map clearly reveals the position of the haem Fe-atom sites, thus paving the way for SAD structure determination.

1. Introduction

Haemoglobins and related haem proteins such as myoglobins, generally referred to as 'globins', have been widely identified not only in animals, but also in plants, fungi and bacteria (Hardison, 1998). Most globins help supply O_2 to the aerobic metabolism of the respiratory chain (Antonini & Brunori, 1971; Perutz, 1990; Wittenberg, 1992; Brunori, 1999; Merx *et al.*, 2002), although in some cases they have been shown to display enzymatic functions (Minning *et al.*, 1999; Brunori, 2001; Flögel *et al.*, 2001). Four types of globins, differing in fine structure, tissue distribution and likely in function, have been discovered in man and other vertebrates: haemoglobin (Hb), myoglobin (Mb), neuroglobin (Ngb) and cytoglobin (Cygb). Hb and Mb have been extensively characterized in terms of structure, function and evolution (Antonini & Brunori, 1971; Wittenberg & Wittenberg, 1989; Perutz, 1990; Hardison, 1998; Brunori, 1999, 2001; Merx *et al.*, 2002). Very recently, Ngb, which is predominantly expressed in the nervous system, and Cygb, which is present in almost all tissue types (alternatively known as histoglobin), have been identified as members of the Hb superfamily (Burmester *et al.*, 2000, 2002; Pesce *et al.*, 2002; Trent & Hargrove, 2002).

Cygb has been recognized in the EST databases of man, mouse and zebrafish (Burmester *et al.*, 2002; Trent & Hargrove, 2002). The rat homologue had been previously identified (as stellate cell-activated protein, STAP) in a proteomics approach that showed up-regulation of a haem-containing protein in the stellate cells of fibrotic liver (Kawada *et al.*, 2001). The size of mammalian (190 amino acids) and fish (174 amino acids) Cygb exceeds that of vertebrate Mbs and Hbs, which typically comprise 140–150 residues (Burmester *et al.*, 2002; Pesce *et al.*, 2002; Trent & Hargrove, 2002). The structural differences relative to Hb and Mb are mainly located in the unusually long N- and C-terminal regions, whereas structural features of the classical globin fold can be recognized in the Cygb core region, the amino-acid sequence of which suggests conservation of the classical three-over-three α-helical sandwich fold (Bolognesi *et al.*, 1997; Neuwald *et al.*, 1997; Hardison, 1998; Burmester *et al.*, 2002; Pesce *et al.*, 2002; Trent & Hargrove, 2002). Thus, the key residues PheCD1, HisE7 and HisF8, which are important for structure and function in respiratory globins, are conserved in Cygbs. Sequence conservation among mammalian Cygbs is extremely high: mouse and human Cygb differ by only 4% of their amino-acid sequences.

Cygb belongs to a newly discovered class of vertebrate hexacoordinated Hbs (Burmester *et al.*, 2002; Trent & Hargrove, 2002), where in the absence of a haem exogenous ligand (*e.g.* O_2 or CO) the haem Fe atom is coordinated to the haem distal HisE7 residue. Kinetics and thermodynamics for O_2 and CO binding to Cygb follow the minimum three-state mechanism, accounting for the endogenous ligand dissociation to allow oxygenation and carbonylation. Thus, ligand affinity for Cygb (P_{50} = 73.3 and 4.53 Pa for O_2 and CO, respectively) is comparable with that displayed by Mb (P_{50} = 86.7 and 22.7 Pa for O_2 and CO, respectively; Trent & Hargrove, 2002; Pesce *et al.*, 2002).

Cygb has been proposed to play an Mb-like function in O_2 metabolism, facilitating O_2 diffusion to the mitochondria (Burmester *et al.*, 2000, 2002; Moens & Dewilde, 2000). Moreover, Cygb may act as an NADH oxidase, facilitating the glycolytic production of ATP

under semi-anaerobic conditions, as proposed for the hexacoordinated non-symbiotic Hbs of plants, *e.g.* maize (Sowa *et al.*, 1998). Cygb may also function as an O_2 sensor, participating in a signal transduction pathway that modulates the activities of regulatory proteins in response to changes in O_2 concentration. As reported for Mb (Ascenzi *et al.*, 2001; Brunori, 2001; Flögel *et al.*, 2001), flavohaemoglobins (Liu *et al.*, 2000), truncated Hbs (Ouellet *et al.*, 2002) and *Ascaris* Hb (Minning *et al.*, 1999), Cygb might be involved in NO/O_2 chemistry. Finally, Cygb displays a weak catalase activity, the physiological relevance of which is unclear as this property is shared by many different globins (Kawada *et al.*, 2001).

In the context of a detailed structural and functional characterization of this unexpected and new member of the Hb superfamily, our laboratories have produced orthorhombic crystals of human Cygb, which are described here together with their preliminary crystallographic characterization.

2. Materials and methods

2.1. Cytoglobin expression and purification

The expression plasmid (the human double mutant of Cygb cDNA, having both cysteines replaced by serines, in pET3a; Burmester *et al.*, 2002) was transformed into *Escherichia coli* strain BL21(DE3)pLysS. The cells were grown at 298 K in TB medium containing 200 µg ml^{-1} ampicillin and 30 µg ml^{-1} chloramphenicol. The culture was induced at $A_{550} = 0.8$ by addition of isopropyl-1-thio-D-galactopyranoside to a final concentration of 0.04 m*M* and expression was continued overnight. Overexpression of Cygb led to production of inclusion bodies. Cells were harvested and resuspended in 50 m*M* Tris–HCl pH 7.5, 0.5 *M* NaCl, 1 m*M* EDTA, 1 m*M* PMSF and 5 m*M* DTT. One-tenth of a volume of a buffer solution containing 10% Triton X-100, 10% deoxycholic acid, 500 m*M* Tris–HCl pH 7.5 and 20 m*M* EDTA was added. The cells were then exposed to three freeze–thaw steps and sonicated until completely lysed. The inclusion bodies were sedimented by centrifugation at 10 000*g* for 10 min. The pellet (inclusion bodies) was washed four times in 1% Triton X-100, 1 m*M* EDTA and 50 m*M* Tris–HCl pH 7.5. The inclusion bodies were then solubilized in 6 *M* guanidinium chloride, 50 m*M* Tris–HCl pH 7.5 and 1% β-mercaptoethanol and left for 1 h on ice. Insoluble material was removed by centrifugation at 10 000*g* for 10 min. Cygb was refolded as follows. Haemin was dissolved in 0.1 *M* NaOH and 1/10 diluted with 50 m*M* Tris–HCl pH 7.5. A 1.4 *M* excess of haemin solution was added to the solubilized proteins. This solution was then dialyzed against 5 m*M* Tris–HCl pH 8.5. Precipitate formed during dialysis was removed by centrifugation at 10 000*g* for 10 min. The refolded Cygb solution was concentrated by Amicon filtration (PM10) and passed through a Sephacryl S200 column. The pure Cygb fractions were pooled, concentrated and stored at 253 K.

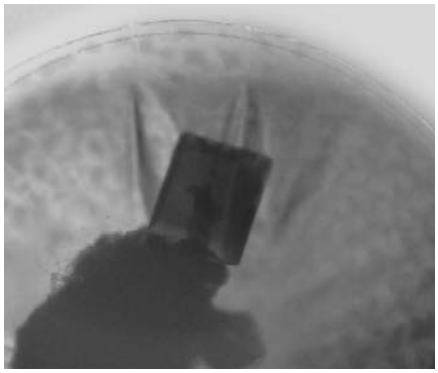

Figure 1
View of an orthorhombic Cygb crystal grown at 277 K. The crystal dimensions are ~0.3 × 0.5 × 0.5 mm.

2.2. Crystallization

Preliminary crystallization conditions were screened using Hampton Sparse Matrix Crystal Screens 1 and 2 (Hampton Research, Laguna Hills, California, USA). After optimization (of pH and precipitant concentration), the following conditions were adopted for reproducible production of Cygb crystals using the hanging-drop vapour-diffusion setup. The reservoir solution contained 20%(*w*/*v*) PEG 4000, 200 m*M* NaCl, 0.05 *M* sodium acetate, 0.01 *M* potassium ferricyanide and 1 m*M* KCN, pH 4.0–5.0; the droplet was composed of 1 µl protein solution (44 mg ml^{-1}) and 1 µl reservoir solution. Under these conditions, Cygb crystals grew at 277 and 294 K as orthorhombic prisms of considerable size (Fig. 1) in about two months. After recovery from the growth chambers, Cygb crystals were transferred into a stabilizing solution which had the same composition as the reservoir solution but contained 30%(*w*/*v*) PEG. For cryoprotection during data collection, the same stabilizing solution was supplemented with 20%(*v*/*v*) glycerol.

2.3. Data collection and SAD phasing

X-ray diffraction data were collected at the ELETTRA synchrotron source (Trieste, Italy), beamline XRD1 (provided with a MAR CCD detector) at 100 K using a crystal grown at 277 K. The crystal diffracted to a maximum resolution of 2.6 Å. To optimize the haem Fe anomalous signal for single-wavelength anomalous diffraction phasing (SAD), the source wavelength was set to 1.720 Å. The data were collected in dose mode (the dose was 6435 kHz s), with a rotation range of 1° per frame. To optimize the signal-to-noise ratio, the total number of images collected was 530. The data were processed, scaled and merged using the *HKL* package (Otwinowski & Minor, 1997). Data-collection statistics are reported in Table 1. Analysis of the Cygb self-rotation function was performed using programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

Table 1
Data-collection statistics for Cygb.

Outer-shell statistics (2.69–2.60 Å) are shown in parentheses.

Source	ELETTRA (beamline XRD1)
Wavelength (Å)	1.720
Resolution limits (Å)	40.0–2.6
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 46.8, b = 73.1, c = 98.9
Mosaicity (°)	0.60
Total reflections	498807
Unique reflections	19995
Completeness (%)	99.5 (96.6)
Redundancy	11 (10)
$R_{\rm merge}$† (%)	5.7 (15.9)
Average $I/\sigma(I)$	35 (11)

† $R_{\rm merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$.

3. Results and discussion

Analysis of the symmetry and systematic absences in the recorded diffraction pattern indicates that Cygb crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 46.8, b = 73.1, c = 98.9 Å. Calculation of the crystal-packing parameter indicates that two Cygb molecules can be accommodated in the asymmetric unit ($V_{\rm M} = 2.05$ Å^3 Da^{-1}), with a solvent content of 40% (Matthews, 1968). Accordingly, the protein self-rotation function, calculated in the 40.0–2.6 Å resolution range, allowed us to recognize promptly that the second highest peak (after identity) in the self-rotation space was located in the $\chi = 180°$ section, indicating the presence of a local twofold axis lying in the unit-cell *yz* plane at an angle of about 30° relative to the *y* axis.

The reduced diffraction data were subsequently used to calculate an anomalous difference Patterson map using *SOLVE* (Terwilliger & Berendzen, 1999), which permitted prompt location and refinement

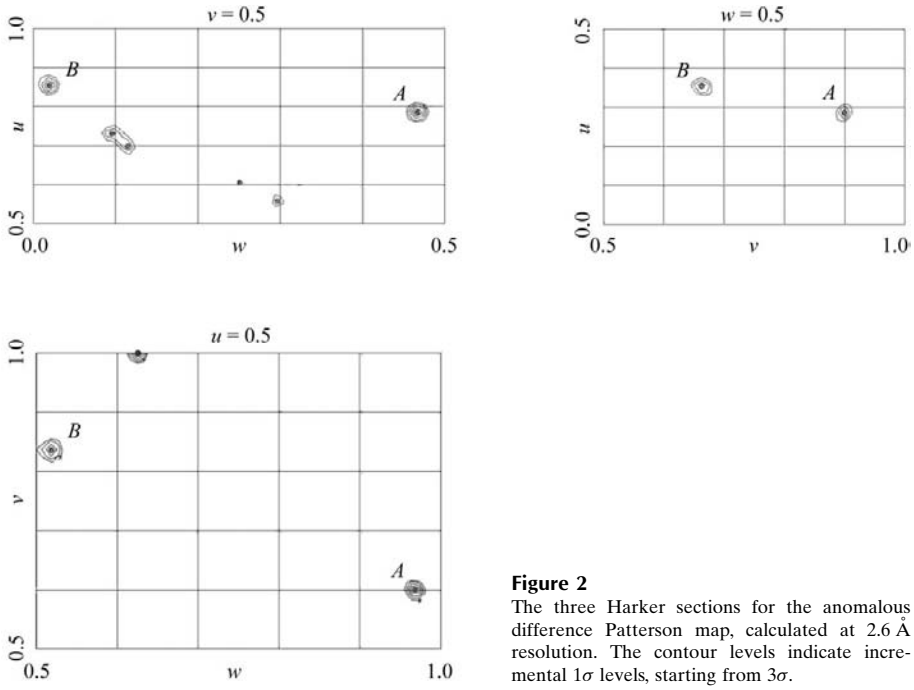

Figure 2
The three Harker sections for the anomalous difference Patterson map, calculated at 2.6 Å resolution. The contour levels indicate incremental 1σ levels, starting from 3σ.

of two Fe-atom sites per asymmetric unit (Fig. 2). Following this, initial phasing was performed using Bayesian correlated phasing, obtaining a final figure of merit of 37%. Maximum-likelihood density modification of the initial electron-density map was then performed using *RESOLVE* (Terwilliger & Berendzen, 1999) and improved using non-crystallographic symmetry averaging over the two molecules in the asymmetric unit with the program *DM* (Cowtan, 1994). Some main elements of the protein secondary structure could be recognized in both independent molecules for the Cygb core regions surrounding the haem and were automatically built. Manual model building to complete the model together with refinement and analysis of the structure are in progress.

We are grateful to Dr Marco Nardini for helpful discussion. This work was supported by grants from the EU Project 'Neuroglobin and survival of the neuron' (QLG3-CT-2002-01548), by the Italian CNR project 'Genomica Funzionale', by ASI grant I/R/294/02 and by German DFG grants Ha2103/3 and Bu956/5. SD is a postdoctoral fellow from the FWO (fund for Scientific Research-Flanders). MB is grateful to Istituto G. Gaslini, Genova for continuous support.

References

Antonini, E. & Brunori, M. (1971). *Hemoglobin and Myoglobin in their Reaction with Ligands.* Amsterdam: North Holland.
Ascenzi, P., Salvati, L. & Brunori, M. (2001). *FEBS Lett.* **501**, 103–105.
Bolognesi, M., Bordo, D., Rizzi, M., Tarricone, C. & Ascenzi, P. (1997). *Prog. Biophys. Mol. Biol.* **68**, 29–68.
Brunori, M. (1999). *Trends. Biochem. Sci.* **24**, 158–161.
Brunori, M. (2001). *Trends Biochem. Sci.* **26**, 209–210.
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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Flögel, U., Merx, M. W., Gödecke, A., Decking, U. K. & Schrader, J. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 735–740.
Hardison, R. C. (1998). *J. Exp. Biol.* **201**, 1099–1117.
Kawada, N., Kristensen, D. B., Asahina, K., Nakatani, K., Minamiyama, Y., Seki, S. & Yoshizato, K. (2001). *J. Biol. Chem.* **276**, 25318–25323.
Liu, L., Zeng, M., Hausladen, A., Heitman, J. & Stamler, J. S. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 4672–4676.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merx, M. W., Flögel, U., Stumpe, T., Gödecke, A., Decking, U. K. & Schrader, J. (2002). *FASEB J.* **15**, 1077–1079.
Minning, D. M., Gow, A. J., Bonaventura, J., Braun, R., Dewhirst, M., Goldberg, D. E. & Stamler, J. S. (1999). *Nature (London)*, **401**, 497–502.
Moens, L. & Dewilde, S. (2000). *Nature (London)*, **407**, 461–462.
Neuwald, A. F., Liu, J. S., Lipman, D. J. & Lawrence, C. E. (1997). *Nucleic Acids Res.* **25**, 1665–1677.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ouellet, H., Ouellet, Y., Richard, C., Labarre, M., Wittenberg, B., Wittenberg, J. & Guertin, M. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 5902–5907.
Perutz, M. F. (1990). *Annu. Rev. Physiol.* **52**, 1–25.
Pesce, A., Bolognesi, M., Bocedi, A., Ascenzi, P., Dewilde, S., Moens, L., Hankeln, T. & Burmester, T. (2002). *EMBO Rep.* **3**, 1146–1151.
Sowa, A. W., Duff, S. M. G., Guy, P. A. & Hill, R. D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 10317–10321.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Trent, J. T. III & Hargrove, M. S. (2002). *J. Biol. Chem.* **277**, 19538–19545.
Wittenberg, B. A. & Wittenberg, J. B. (1989). *Annu. Rev. Physiol.* **51**, 857–878.
Wittenberg, J. B. (1992). *Adv. Comput. Environ. Physiol.* **13**, 60–85.