EMBO reports

scientific report

Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity

Marco Nardini[1], Alessandra Pesce[2], Liesbet Thijs[3], Jennifer A. Saito[4], Sylvia Dewilde[3], Maqsudul Alam[4], Paolo Ascenzi[5], Massimiliano Coletta[6], Chiara Ciaccio[6], Luc Moens[3] & Martino Bolognesi[1+]

[1]Department of Biomolecular Sciences and Biotechnology, CNR-INFM, University of Milano, Milano, Italy, [2]Department of Physics, CNR-INFM and CEBR, University of Genova, Genova, Italy, [3]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, [4]Department of Microbiology, University of Hawaii, Honolulu, Hawaii, USA, [5]Department of Biology and Interdepartmental Laboratory for Electron Microscopy, University Roma Tre, Roma, Italy, and [6]Department of Experimental Medicine and Biochemical Sciences, University of Roma 'Tor Vergata', Roma, Italy

The structural adaptability of the globin fold has been highlighted by the recent discovery of the 2-on-2 haemoglobins, of neuroglobin and cytoglobin. Protoglobin from *Methanosarcina acetivorans* C2A—a strictly anaerobic methanogenic Archaea—is, to the best of our knowledge, the latest entry adding new variability and functional complexity to the haemoglobin (Hb) superfamily. Here, we report the 1.3 Å crystal structure of oxygenated *M. acetivorans* protoglobin, together with the first insight into its ligand-binding properties. We show that, contrary to all known globins, protoglobin-specific loops and an amino-terminal extension completely bury the haem within the protein matrix. Access of O_2, CO and NO to the haem is granted by the protoglobin-specific apolar tunnels reaching the haem distal site from locations at the B/G and B/E helix interfaces. Functionally, *M. acetivorans* dimeric protoglobin shows a selectivity ratio for O_2/CO binding to the haem that favours O_2 ligation and anticooperativity in ligand binding. Both properties are exceptional within the Hb superfamily.

Keywords: archaea protein; haemoprotein structure; methanogenesis; protein matrix tunnels; protoglobin

EMBO reports (2008) **9**, 157–163. doi:10.1038/sj.embor.7401153

[1]Department of Biomolecular Sciences and Biotechnology, CNR-INFM, University of Milano, Via Celoria 26, Milano 20133, Italy
[2]Department of Physics, CNR-INFM and CEBR, University of Genova, Via Dodecaneso, 33, Genova 16146, Italy
[3]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp B-2610, Belgium
[4]Department of Microbiology, University of Hawaii, Snyder Hall 207, 2538 The Mall, Honolulu, Hawaii 96822, USA
[5]Department of Biology and Interdepartmental Laboratory for Electron Microscopy, University Roma Tre, Viale Guglielmo Marconi 446, Roma 00146, Italy
[6]Department of Experimental Medicine and Biochemical Sciences, University of Roma 'Tor Vergata', Via di Tor Vergata 135, Roma 00133, Italy
[+]Corresponding author. Tel: +39 02 50314893; Fax: +39 02 50314895;
E-mail: martino.bolognesi@unimi.it

Received 30 July 2007; revised 12 November 2007; accepted 19 November 2007; published online 11 January 2008

INTRODUCTION

Methanogenesis has a crucial role in the global carbon cycle and contributes significantly to global warming; most methane in nature is produced from acetate (Galagan *et al*, 2002). *Methanosarcina acetivorans* C2A—a strictly anaerobic non-motile Archaea, in contrast to most methanogens—can exploit acetate, methanol, CO_2 and CO as carbon sources for methanogenesis. Methane production in this Archaea occurs simultaneously with the formation of a proton gradient that is essential for energy harvesting (Rother & Metcalf, 2004; Lessner *et al*, 2006). The ability to convert CO to methane might indicate a very ancient origin for such metabolic pathways (Ferry & House, 2006).

Despite its strict anaerobic nature, the genome of *M. acetivorans* has genes that can be related to O_2 metabolism; among these, an open reading frame encodes for a 'protoglobin' (Pgb; NP_617780). Pgbs are single-domain haem proteins of approximately 195 amino acids, which are related to the amino-terminal domain of archaeal and bacterial globin-coupled sensor proteins (GCS; Hou *et al*, 2001; Freitas *et al*, 2004; Vinogradov *et al*, 2006, 2007). Sequence comparisons indicate that Pgbs, despite their 30–35% larger size, are structurally related to single-chain Hbs (composed of approximately 150 amino acids, folded into a 3-on-3 α-helical sandwich, with 12–16% residue identity to Pgbs) and to the haem-based aerotaxis transducer sensor domain of *Bacillus subtilis* GCS (*Bs*HemAT). Pgbs bind O_2, CO and NO reversibly *in vitro*. Although functional and evolutionary issues are openly debated (Freitas *et al*, 2004; Vinogradov *et al*, 2006, 2007), Pgb might facilitate O_2 detoxification *in vivo* by promoting electron transfer to O_2, or might act as CO sensor/supplier in methanogenesis. Our study provides the first information, to our knowledge, of the three-dimensional structure of Pgb and provides preliminary results on the Pgb ligand-binding properties (versus O_2 and CO) that underline an almost unique behaviour within the haemoglobin (Hb) superfamily.

RESULTS AND DISCUSSION

The crystal structures of a Pgb (oxygenated Pgb from *M. acetivorans*, *Ma*Pgb-O_2, solved at 1.3 Å resolution, one

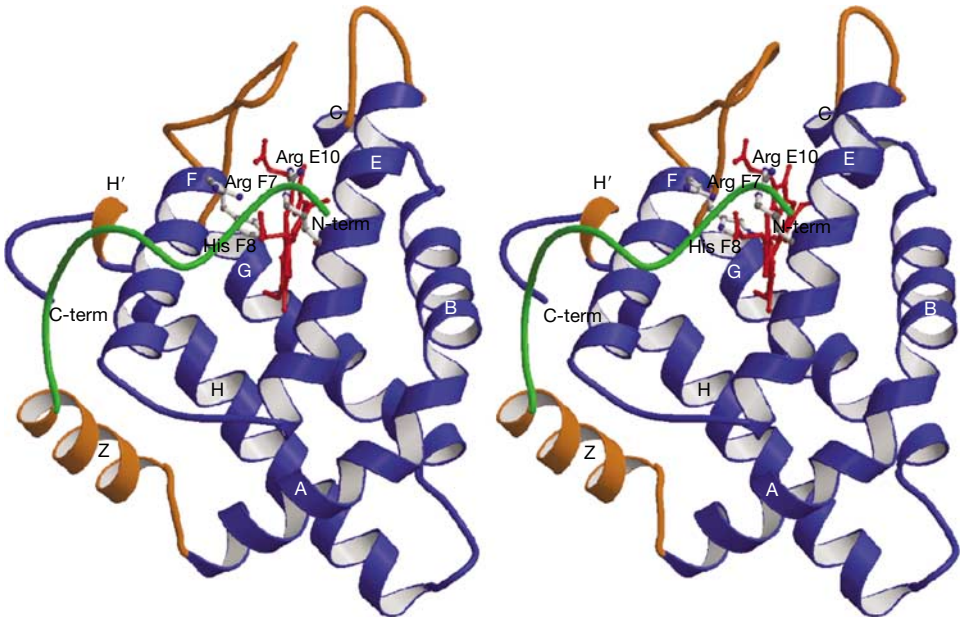

Fig 1 | Stereo view of the *Methanosarcina acetivorans* protoglobin fold. The figure highlights the elements with secondary structure (blue; labels A–H′). The main structural elements of the protein that are specific to *Ma*Pgb (relative to 3-on-3 Hbs) are shown in orange, except for the first 20 amino-terminal residues that are in green. Notice the N-terminal region, the CE and FG loops that bury the haem (edge on and red colour, salt-linked with Arg(E10) and Arg(F7)) and prevent access of small ligands to the haem distal cavity. The proximal His(F8) residue is shown on the left hand side of the haem. Hbs, haemoglobins; *Ma*Pgb, *Methanosarcina acetivorans* protoglobin.

molecule per asymmetric unit, and the ligand-free form of ferric *Ma*Pgb, solved at 2.6 Å resolution, eight molecules per asymmetric unit; Table 1) highlight a protein fold consisting of nine main helices (labelled Z, A, B, C, E, F, G, H and H′), which are partly related to those identified in Hb (Bolognesi *et al*, 1997) and in GCS (Zhang & Phillips, 2003; Fig 1; supplementary Fig S1 online). The Z-helix precedes the globin-fold conserved A-helix; it is preceded by 20 amino acids building an N-terminal loop held, in both structures, next to the haem propionates by hydrogen bonds that link Pro 7, Gly 8, Tyr 9, Thr 10, Ala 18 and Phe 20 to residues of the E- and F-helices in the protein α-helical core. The haem propionates are thus solvent inaccessible and stabilized by intramolecular salt-bridges to Arg(E10)92 and Arg(F7)119 (Fig 1). The interaction between the N-terminal loop and the protein core is not mediated by water molecules, resulting in a rigid loop structure (atomic B-factors in the first 20 *Ma*Pgb-O_2 residues average at 13.5 $Å^2$). In addition, we note that conservation of an amino-acid sequence motif (IPGYxYG) in the N-terminal loop seems very strong among known Pgbs, whereas no sequence correlation is found with the N-terminal region of GCS (supplementary Fig S2 online).

The haem group crevice of *Ma*Pgb-O_2 is built by the distal B-, C- and E-helices, and by the proximal F-helix; His(F8)120 is the proximal residue coordinating the haem-Fe atom. Contrary to all known Hb structures (in which approximately 30% of the haem surface is solvent accessible), the haem group of *Ma*Pgb-O_2 is fully buried within the crevice. Such a unique structural feature is related to the conformation of the 1–20 N-terminal segment, and to the extended CE and FG loops (residues 73–81 and 122–137, respectively). All such haem shielding protein segments are conserved in known Pgbs (supplementary Fig S2 online).

The diatomic ligand diffusion path to the haem pocket through the E7-gate first reported for myoglobin (Mb; Bolognesi *et al*, 1982) is thus precluded, whereas *Ma*Pgb-O_2 shows alternate paths. A straight apolar protein matrix tunnel (about 7 Å in diameter) connecting the protein surface to the haem distal cavity is located between the B- and G-helices (tunnel 1; Fig 2). A second straight opening on the haem distal cavity is nestled between the B- and the E-helices (about 5 Å diameter), and is partly defined by Tyr(B10)61 (tunnel 2; Fig 2). Both tunnels host one water molecule at their solvent side aperture. In addition, a core cavity of about 75 $Å^3$ is located between the distal and proximal haem sides; the cavity hosts four mutually hydrogen-bonded water molecules. All the residues lining the protein tunnels and the inner cavity are conserved in known Pgbs (supplementary Fig S2 online), which is consistent with their implication in diatomic ligand diffusion to and from the haem, multi-ligand storage and/or (pseudo-)enzymatic reactivity. Such functional roles would rely on structural principles that are entirely different from those shown for Xe cavities in Mb and in truncated Hbs (Brunori & Gibson, 2001; Milani *et al*, 2005).

In *Ma*Pgb-O_2, the haem-bound O_2 molecule is not stabilized by hydrogen bonds to the protein. The residue closest to O_2 (Phe(E11)93 at 3 Å) is affected by conformational disorder (Fig 3). By contrast, Phe(E11)93 adopts a defined conformation in the absence of the haem ligand, as shown by the crystal structure of the ferric ligand-free *Ma*Pgb. In this case, the fixed conformation of Phe(E11)93 triggers a rotation of approximately 45° of Tyr(B10)61, giving rise to a wide distal site cavity that is directly connected to the aforementioned apolar tunnels.

The high resolution of the *Ma*Pgb-O_2 structure provides unequivocal evidence of substantial distortion in the porphyrin

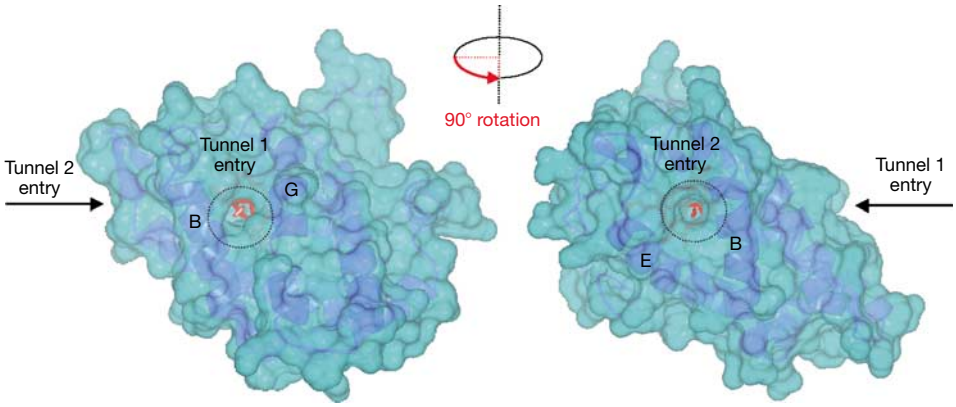

Fig 2 | Mono views of 'tunnel 1' and 'tunnel 2' access sites in *Methanosarcina acetivorans* protoglobin. The protein molecular surface (as defined by a 1.4 Å radius probe) is shown in a semitransparent cyan colour, allowing the underlying helices (blue ribbons) to be seen. Helices flanking the tunnel entries are labelled. The haem group (seen through the tunnel apertures highlighted in the figure by dotted circles) is shown as red sticks. The protein is correctly oriented in both images, to bring each tunnel in the direction of sight; arrows show the location of the tunnel that is not in view, for each image.

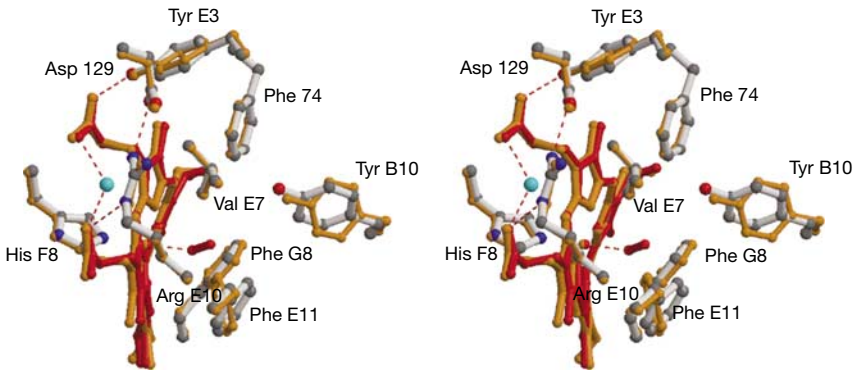

Fig 3 | Stereo view of *Methanosarcina acetivorans* protoglobin distal site organization. The figure shows the main residues building the haem distal site cavity, hosting an O_2 molecule (red), and the conformations adopted in *Ma*Pgb-O_2 (white bonds and red haem) and in the ligand-free *Ma*Pgb (orange). When appropriate, topological positions of the residues are indicated. Hydrogen bonds are represented as dashed lines; one water molecule participating to haem propionates stabilization is shown as a cyan sphere. *Ma*Pgb-O_2, oxygenated protoglobin from *Methanosarcina acetivorans*.

ring system. The haem-Fe atom falls slightly out of the haem plane (−0.13 Å) towards the proximal site; however, out-of-plane deviations of only ± 0.5–0.6 Å affect the four-haem pyrrole rings. Specific *Ma*Pgb–haem contacts include residue–pyrrole π–π interactions, hydrogen bonds and salt bridges at the propionates, and interaction with a buried water molecule. As deviation of porphyrins from planarity is energetically unfavourable, the marked haem distortion indicates a proactive role of *Ma*Pgb in modulating haem reactivity (for example, chemical and photophysical properties, axial ligand affinity, O_2 on/off rates, redox potentials, transition dipoles and energies; Shelnutt *et al*, 1998), providing a mechanism for differentiating between CO and O_2 binding. Carbonylation and oxygenation reactions in *Ma*Pgb are reminiscent of those in the *Bs*HemAT sensor domain (Zhang *et al*, 2005), but are strikingly different from those found in known globins. In fact, *Ma*Pgb shows multiphasic (de)carbonylation and (de)oxygenation kinetics, as well as anticooperative ligand-binding equilibria ($n_{50} \sim 0.5$; see supplementary Fig S3 online). Table 2 shows the kinetic and thermodynamic parameters for O_2 and CO binding to *Ma*Pgb compared with those of other prototypical haem proteins. The comparison highlights that (i) the functional behaviour for both ligands is biphasic, indicating the presence of two binding sites in *Ma*Pgb with different functional properties, (ii) the overall O_2 affinity for *Ma*Pgb ($P_{50} \sim 2.5$ mmHg) is higher than that of CO ($P_{50} \sim 10$ mmHg), a ligand-binding behaviour that is exceptional within the Hb superfamily, and (iii) such a binding property is exclusively referable to unusually fast CO dissociation rate constants (see Table 2; supplementary Fig S4 online).

Gel filtration experiments on the expressed protein show that *Ma*Pgb elutes as a dimer in solution (data not shown). Accordingly, crystal packing in the C2 crystal form constructs a symmetrical *Ma*Pgb-O_2 homodimer, in which the 2,086 Å^2 association interface is contributed mostly by residues belonging to the G- and H-helices (which build an intermolecular four-helix bundle), to the H′-helix, partly to the Z-helix and to the BC and FG hinges (supplementary Fig S5 online). Furthermore, the ligand-free *Ma*Pgb crystal structure, showing four independent homodimers in the crystallographic asymmetrical unit, highlights a closely similar quaternary assembly. Comparison of the quaternary

Table 1 | Data collection, phasing (MAD: Fe) and crystallographic refinement statistics for *Ma*Pgb-O_2 and unliganded *Ma*Pgb

		Crystal *Ma*Pgb-O_2		**Crystal *Ma*Pgb**
Data collection				
Space group		C2		$P2_1$
Cell dimensions:				
a, *b*, *c* (Å)		80.1, 49.3, 51.5		101.9, 48.5, 161.1
α, β, γ (deg)		90, 92.9, 90		90, 98.4, 90
	Peak	*Inflection*	*Remote*	
Wavelength (Å)	1.738	1.740	0.954	0.931
Resolution (Å)	51.4–1.7	51.4–1.7	42.0–1.3	58.4–2.6
No reflections	179,215	81,427	338,891	115,927
Unique reflections	19,890	18,620	48,188	47,411
R_{merge}	0.045 (0.251)	0.104 (0.112)	0.048 (0.215)	0.087 (0.350)
$I/\sigma(I)$	35.6 (4.5)	11.6 (5.5)	24.7 (7.5)	9.6 (2.7)
Completeness (%)	90.3 (46.8)	84.4 (44.8)	97.8 (99.8)	97.3 (98.0)
Multiplicity	9.0 (5.7)	4.4 (3.1)	7.0 (7.1)	2.4 (2.4)
Refinement				
R_{factor}/R_{free}		16.0/19.0		20.5/26.3
No. of residues/protein atoms		190 (6–195)/1,603		$8 \times 190/12{,}728$
No. of haem groups		1		8
No. of O_2 molecules		1		–
No. of phosphate ions		2		–
No. of glycerol molecules		1		–
No. of water molecules		265		238
B-factors (Å²)				
Protein		15.3		30.4
Haem group		8.4		19.6
Ligand/ion		19.7		–
Water molecules		32.6		28.7
R.m.s. deviation from ideality				
Bond lengths (Å)		0.010		0.007
Bond angles (deg)		1.3		1.2

Values in parentheses are for the highest-resolution shell.
MAD, multiple wavelength anomalous dispersion; *Ma*Pgb-O_2, oxygenated protoglobin from *Methanosarcina acetivorans*.

structures shows that small but perceptible conformational changes occur on ligand binding to *Ma*Pgb (overall r.m.s. deviation between *Ma*Pgb-O_2 and the ligand-free dimers of approximately 0.8 Å), coupled to acquisition of exact twofold symmetry in the dimer.

Overall, the tertiary and quaternary structures of *Ma*Pgb are similar to those of the *Bs*HemAT sensor domain (Zhang & Phillips, 2003; supplementary Fig S1 online), with which *Ma*Pgb shares approximately 10% sequence identity. The r.m.s. deviation between the isolated chains of *Ma*Pgb and *Bs*HemAT is 1.7 Å (limited to 142 C_α pairs). Major structural differences are evident at the 20 N-terminal residues, running in opposite directions, at the CE loop region (where *Ma*Pgb shows a 6-residue insertion), at the FG loop (12-residue insertion in *Ma*Pgb), and at the carboxy-terminal H′-helix (not present in *Bs*HemAT; supplementary Figs S1,S2 online). Most importantly, such structural differences render the haem distal site of *Bs*HemAT accessible to solvent from the propionates edge, whereas specific mutations (supplementary Fig S2 online) and local shifts of the B-, E- and G-helices prevent the formation of the apolar tunnels observed in *Ma*Pgb.

The globin domain has been adopted by nature to host the haem in an increasing number of successful engineering

Table 2 | Values of kinetic and thermodynamic parameters for O_2 and CO binding to *Ma*Pgb and prototypical haem-proteins

Protein	O_2*			CO*			
	k_{on} ($M^{-1}\,s^{-1}$)	k_{off} (s^{-1})	K (M^{-1})	h_{on} ($M^{-1}\,s^{-1}$)	h_{off} (s^{-1})	H (M^{-1})	K/H
Sperm whale Mb†	1.7×10^{7}	1.6×10^{1}	*1.1×10^{6}*	5.1×10^{5}	1.9×10^{-2}	*2.7×10^{7}*	*4.1×10^{-2}*
Vitreoscilla Hb‡	2.0×10^{8}	4.2	*4.8×10^{7}*	ND	ND	ND	ND
		1.5×10^{-1}	*1.3×10^{9}*				
Bacillus subtilis flavoHb§	4.4×10^{7}	7.0	*6.3×10^{6}*	7.0×10^{6}	3.6×10^{-1}	*1.9×10^{7}*	*3.3×10^{-1}*
	2.8×10^{6}	3.0×10^{-1}	*9.3×10^{6}*	5.0×10^{5}		*1.4×10^{6}*	*6.6*
Bacillus subtilis trHbO‖	1.4×10^{7}	3.5×10^{-3}	4.0×10^{9}	2.2×10^{5}	4.6×10^{-4}	4.8×10^{8}	*8.3*
Bacillus subtilis HemAT¶	1.9×10^{7}	1.9×10^{3}	1.0×10^{4}	3.4×10^{5}	6.7×10^{-2}	5.1×10^{6}	*1.9×10^{-3}*
		8.7×10^{1}	2.2×10^{5}				*4.3×10^{-2}*
*Ma*Pgb#	*1.7×10^{4}***	9.4×10^{-3}***	1.9×10^{6}***	5.0×10^{6}***	3.9×10^{1}***	1.3×10^{5}***	*1.5×10^{1}***
	3.2×10^{3}††	9.2×10^{-2}††	3.4×10^{4}††	2.4×10^{6}††	1.2††	2.0×10^{6}††	*1.7×10^{-2}*††

ND, not determined.
*Values of kinetic and thermodynamic parameters in italic have been calculated according to the following equations: $K = k_{on}/k_{off}$ and $H = h_{on}/h_{off}$. †pH = 7.0 and $T = 20.0\,^{\circ}C$ (Smerdon *et al*, 1995). ‡pH = 7.0 and $T = 20.0\,^{\circ}C$ (Giangiacomo *et al*, 2001). §pH = 7.0 and $T = 20.0\,^{\circ}C$. Oxygenation and deoxygenation kinetics are biphasic. Carbonylation kinetics and thermodynamics are biphasic and anticooperative, respectively (Farrés *et al*, 2005). ‖pH = 7.0 and $T = 20.0\,^{\circ}C$ (Giangiacomo *et al*, 2005). ¶pH = 7.0 and $T = 25.0\,^{\circ}C$. Deoxygenation kinetics is biphasic and oxygenation thermodynamics is anticooperative (Zhang *et al*, 2005). #pH = 7.0 and $T = 20.0\,^{\circ}C$. Ligand-binding kinetics is biphasic and thermodynamics is anticooperative (see Scheme SI). The s.d. of values of the experimentally determined parameters is ± 10% (this study; see supplementary information online).
**Ligand binding to *Ma*Pgb conforms to $K_1 = k_{on1}/k_{off1}$ and $H_1 = h_{on1}/h_{off1}$ (see Scheme SI). ††Ligand binding to *Ma*Pgb-L conforms to $K_2 = k_{on2}/k_{off2}$ and $H_2 = h_{on2}/h_{off1}$ (see Scheme SI).
Hb, haemoglobin; *Ma*Pgb, *Methanosarcina acetivotans* prologlobin; Mb, myoglobin.

experiments. These include single domain (3-on-3) Hbs, flavohaemoglobins, truncated (2-on-2) Hbs and GCSs (Vinogradov *et al*, 2007). Structural modulation of the 3-on-3 fold in Pgb translates into new access routes to the haem, into unique modulation of haem structure and reactivity, and into a specific quaternary assembly. From a functional viewpoint, such structural properties code for a strikingly modified O_2/CO selectivity ratio and for anticooperativity in ligand binding. Given the unusual occurrence of such properties within the whole Hb superfamily, a specific link to the metabolic behaviour of *M. acetivorans* seems plausible. Taken together, and contrary to earlier suggestions (Vinogradov *et al*, 2007), these results support the idea that Pgb is the evolutionary archetype of GCSs, but not of all globins.

METHODS

Expression, cloning and mutagenesis. The *M. acetivorans* strain was purchased from American Type Culture Collection. Genomic DNA was isolated using a GNOME DNA isolation kit (QBiogene, Morgan Irvine, CA, USA). The gene fragments encoding *Ma*Pgb (codons 1–195) were amplified by PCR. The PCR products were cloned into the PCR4Blunt-TOPO vector (Invitrogen, Carlsbad, CA, USA), followed by subcloning into a pET3a vector. For crystallization purposes, the Cys 101 → Ser mutation was introduced (QuickChange™ site directed mutagenesis, Stratagene, La Jolla, CA, USA). The following primers were used for mutation experiments: 5′-TGGATTCTTGATACCTCCAATCGCAGCTATG-3′ and 5′-CATAGCTGCGATTGGAGGTATCAAGAATCCA-3′.

Purification. *Ma*Pgb was expressed in *Escherichia coli* cells Bl21(DE3)pLysS and collected as described previously (Dewilde *et al*, 2001). Alternatively, the cells were resuspended in 50 mM Tris–HCl (pH 8.0), 5 mM EDTA and 1 mM phenylmethylsulphonyl fluoride. The cells were then exposed to three freeze–thaw steps and sonicated until completely lysed. Inclusion bodies were washed twice with 50 mM Tris–HCl (pH 8.0), 5 mM EDTA and 2% sodium deoxycholate, washed once with pure water and solubilized by incubation in 6 M GuHCl, 100 mM Tris–HCl (pH 8) and 72 mM dithiothreitol (DTT) for 1 h at 22 °C. After centrifugation at 10,700*g* for 20 min at 4 °C, *Ma*Pgb was refolded by adding a 1.4 M excess of haemin. After a incubation period of 10 min at 22 °C, the solution was diluted into four volumes of refolding buffer (100 mM Tris–HCl (pH 8.0), 0.2 M KCl, 0.4 M arginine, 5 mM DTT, 2% glycine) and dialysed at 4 °C against the refolding buffer (Wright & Serpersu, 2004). Final purification was performed by gel filtration using a Sephacryl S200 column equilibrated in 50 mM Tris (pH 8.5), 150 mM NaCl and 0.5 mM EDTA.

Crystallization and data collection. Crystallization of *Ma*Pgb was achieved in hanging-drop setups. The protein solution at 44 mg/ml was equilibrated against a precipitant solution containing 0.4 M monobasic ammonium phosphate (pH 4.2), at 277 K. Large single crystals, grown within few days, were stored in 0.8 M ammonium phosphate, and then transferred to the same solution supplemented with 20% (v/v) glycerol, before cryo-cooling and data collection. The crystals diffract up to 1.3 Å resolution, using synchrotron radiation (beamline ID23-1, ESRF, Grenoble, France), and belong to the centred monoclinic C2 space group, with unit cell parameters $a = 80.1$ Å, $b = 49.3$ Å, $c = 51.5$ Å and $\beta = 92.9\,^{\circ}$. The V_M value is 2.31 $Å^3\,Da^{-1}$, 46.9% solvent content, assuming one *Ma*Pgb molecule per asymmetrical unit. Alternatively, bunches of plates were grown in 1–2 weeks using 10% isopropanol, 20% PEG 4000 and 0.1 M Na HEPES (pH 7.5) as precipitant solution, at 277 K. The crystals were stored in 15% isopropanol, 20% PEG 4000 and 0.1 M Na HEPES (pH 7.5), and transferred to the same solution supplemented with 20% (v/v)

glycerol, before data collection (at 100 K). These crystals diffract up to 2.6 Å resolution (beamline ID14-3, ESRF, Grenoble, France) and belong to the monoclinic space group $P2_1$, with unit cell parameters $a = 101.9$ Å, $b = 48.5$ Å, $c = 161.1$ Å and $\beta = 98.4^\circ$. The V_M value is 2.25 $Å^3\,Da^{-1}$, 45.3% solvent content, assuming eight *Ma*Pgb molecules per asymmetrical unit. All collected data were reduced and scaled using MOSFLM and SCALA (Evans, 1993; Leslie, 2003), and the programmes from the CCP4 suite (CCP4, 1994; Table 1).

Structural determination and refinement. Multiple wavelength anomalous dispersion phases, based on the haem-Fe atom anomalous signal, were determined on the C2 crystal form at 1.3 Å resolution with SOLVE (Terwilliger & Berendzen, 1999). The electron density map was improved by solvent flattening with DM (CCP4, 1994), yielding a figure of merit of 0.79 at 2.0 Å resolution. ARP/wARP (Perrakis *et al*, 1999) was used to extend and refine phases to 1.3 Å resolution, and for automated model building of the main and side chains. The molecular model was checked manually with COOT (Emsley & Cowtan, 2004) and refined using REFMAC (Murshudov *et al*, 1997). At the end of the refinement stages (including anisotropic B-factor refinement), 1 O_2 molecule, 265 water molecules, 2 phosphate ions and 1 glycerol molecule were located through inspection of difference Fourier maps. The final R_{factor} and R_{free} values were 16.0% and 19.0%, respectively (Table 1).

The refined structure of the *Ma*Pgb monomer was used as starting model to solve the $P2_1$ crystal form structure using Phaser (Storoni *et al*, 2004). The rotational and translational searches yielded eight prominent solutions in the 25–2.6 Å resolution range. To reduce model bias, the eight *Ma*Pgb molecules have been rebuilt using the programme ARP/wARP (Perrakis *et al*, 1999) and phases from the molecular replacement solution. At the end of the refinement (Murshudov *et al*, 1997), 238 water molecules were located through the inspection of difference Fourier maps using COOT (Emsley & Cowtan, 2004). The final R_{factor} and R_{free} values were 20.5% and 26.3%, respectively (Table 1). The programmes Procheck and Surfnet (Laskowski *et al*, 1993; Laskowski, 1995) were used to assess stereochemical quality and to explore protein matrix cavities. The programme PISA (Krissinel & Henrick, 2005) was used to identify quaternary assemblies. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (Berman *et al*, 2000) with entry codes 2VEB and 2VEE, and r2VEBsf (1.3 Å resolution) and r2VEEsf (2.6 Å resolution), respectively.

Ligand-binding analysis. Thermodynamics and kinetics of O_2 and CO binding to ferrous *Ma*Pgb were fitted to the minimum two-step mechanism depicted in Scheme SI (Antonini & Brunori, 1971):

$$Ma\text{Pgb} + L \underset{k_{off1}\text{ or }h_{off1}}{\overset{k_{on1}\text{ or }h_{on1}}{\longleftrightarrow}} Ma\text{Pgb-L} + L \underset{k_{off2}\text{ or }h_{off2}}{\overset{k_{on2}\text{ or }h_{on2}}{\longleftrightarrow}} Ma\text{Pgb-L}_2 \qquad \text{(SI)}$$

where *Ma*Pgb is the unligated protein, L is O_2 or CO, *Ma*Pgb-L is the monoligated species, *Ma*Pgb-L_2 is the biligated species, k_{on1} and h_{on1} indicate the second-order rate constants, respectively, for *Ma*Pgb-O_2 and *Ma*Pgb-CO formation, k_{on2} and h_{on2} indicate the second-order rate constants, respectively, for *Ma*Pgb-$(O_2)_2$ and *Ma*Pgb-$(CO)_2$ formation, k_{off1} and h_{off1} indicate the first-order rate constants, respectively, for *Ma*Pgb-O_2 deoxygenation and *Ma*Pgb-CO decarbonylation, and k_{off2} and h_{off2} indicate the first-order rate constants, respectively, for *Ma*Pgb-$(O_2)_2$ deoxygenation and *Ma*Pgb-$(CO)_2$ decarbonylation.

Values of K_1 ($= k_{on1}/k_{off1}$) and K_2 ($= k_{on2}/k_{off2}$) for *Ma*Pgb oxygenation (see Scheme SI) were obtained from the dependence of Y on pO_2 according to equation (S1) (Bolognesi *et al*, 1999):

$$Y = (K_1 \times pO_2 + 2 \times K_2^2 \times pO_2^2)/(2 \times (1 + K_1 \times pO_2 + K_2^2 \times pO_2^2)) \quad \text{(S1)}$$

where Y is the O_2 saturation degree of the haem protein and pO_2 is the O_2 partial pressure. The value of the Hill coefficient (n_{50}) was estimated from the slope of the Hill plot (that is, $\ln(Y/(1-Y))$ versus $\ln pO_2$; see supplementary Fig S3a online) at $\ln(Y/(1-Y)) = 0$ (Antonini & Brunori, 1971).

Values of the first-order rate constant for O_2 dissociation from *Ma*Pgb-$(O_2)_2$ (that is, k_{off2}; see Scheme SI; supplementary Fig S3b online) and from *Ma*Pgb-O_2 (that is, k_{off1}; see Scheme SI; supplementary Fig S3b online) were obtained according to the following equations (Bateman, 1910):

$$[Ma\text{Pgb-}(O_2)_2]_t = [Ma\text{Pgb-}(O_2)_2]_i \times e^{-koff2 \times t} \qquad \text{(S2)}$$

$$[Ma\text{Pgb-}O_2]_t = [Ma\text{Pgb-}(O_2)_2]_i \times (k_{off2} \times ((e^{-koff2 \times t/(koff1 - koff2)}) + (e^{-koff1 \times t/(koff2 - koff1)}))) \qquad \text{(S3)}$$

$$[Ma\text{Pgb}]_t = [Ma\text{Pgb} - (O_2)_2]_i\text{-}([Ma\text{Pgb-}(O_2)_2]_t + [Ma\text{Pgb-}O_2]_t) \qquad \text{(S4)}$$

Both reactions were treated as irreversible first-order process, as sodium dithionite consummates O_2 very quickly (Gibson, 1973).

Values of the second-order rate constants for O_2 binding to *Ma*Pgb (that is, k_{on1}) and to *Ma*Pgb-O_2 (that is, k_{on2}) were calculated according to the following equations (Antonini & Brunori, 1971):

$$k_{on1} = K_1 \times k_{off1} \qquad \text{(S5)}$$

$$k_{on2} = K_2 \times k_{off2} \qquad \text{(S6)}$$

Values of the second-order rate constants for *Ma*Pgb carbonylation (that is, h_{on1} and h_{on2}; see Scheme SI) and of the first-order rate constants for CO dissociation from *Ma*Pgb-$(CO)_2$ (that is, h_{off2} and h_{off1}; see Scheme SI) were obtained from the dependence of the pseudo-first-order rate constants h_1 and h_2 on CO concentration (that is, [CO]; see supplementary Fig S4a online) according to the following equations (Antonini & Brunori, 1971):

$$h_1 = h_{on1} \times [CO] + h_{off1} \qquad \text{(S7)}$$

$$h_2 = h_{on2} \times [CO] + h_{off2} \qquad \text{(S8)}$$

Values of H_1 ($= h_{on1}/h_{off1}$) and H_2 ($= h_{on2}/h_{off2}$) for *Ma*Pgb carbonylation were obtained from the dependence of Y_1 and Y_2 on [CO] (see supplementary Fig S4b online) according to the following equations:

$$Y_1 = [CO]/(H_1^{-1} + [CO]) \qquad \text{(S9)}$$

$$Y_2 = [CO]/(H_2^{-1} + [CO]) \qquad \text{(S10)}$$

where Y_1 is the CO saturation degree of *Ma*Pgb (that is, *Ma*Pgb-CO formation) and Y_2 is the CO saturation degree of *Ma*Pgb-CO (that is, *Ma*Pgb-$(CO)_2$ formation). Values of the Hill coefficient (n_{50}) were estimated from the slope of the Hill plots (that is, $\ln(Y_1/(1-Y_1))$ versus ln [CO] and $\ln(Y_2/(1-Y_2))$ versus ln [CO]) at $\ln(Y_1/(1-Y_1))=0$ and $\ln(Y_2/(1-Y_2))=0$, respectively (Antonini & Brunori, 1971). All results are reported in Table 2 as mean values of at least four experiments (for the experimental details, see supplementary information online). All data were analysed using the MATLAB program (The Math Works Inc, Natick, MA, USA).

Supplementary information is available at *EMBO reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

We thank the Centro Interdisciplinare Materiali e Interfacce Nanostrutturati (University of Milano, to M.B.) for support. This work was supported by grants from the Ministero dell'Istruzione, dell'Universita' e della Ricerca Italy (Fondo per gli Investimenti della Ricerca di Base Project 'Biologia Strutturale', to M.B.) and by the Fund for Scientific Research-Flanders (FWO; to L.M.). S.D. is a postdoctoral researcher of this Fund. M.A. was supported by National Science Foundation grant MCB0446431 and US Army Telemedicine and Advanced Technology Research Center Award no. W81XWH-05-2-0013.

COMPETING INTEREST STATEMENT

The authors declare no competing financial interests.

REFERENCES

Antonini E, Brunori M (1971) *Hemoglobin and Myoglobin in their Reactions with Ligands*. Amsterdam, London: North Holland Publishing Co

Bateman H (1910) Solution of a system of differential equations occurring in the theory of radio-active transformations. *Proc Cambridge Phyl Soc* **15:** 423–427

Berman HM, Westbrook J, Fenz Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* **28:** 235–242

Bolognesi M, Cannillo E, Ascenzi P, Giacometti GM, Merli A, Brunori M (1982) Reactivity of ferric *Aplysia* and sperm whale myoglobins towards imidazole. X-ray and binding study. *J Mol Biol* **158:** 305–315

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

Brunori M, Gibson QH (2001) Cavities packing defects in the structural dynamics of myoglobin. *EMBO Rep* **2:** 674–679

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

Dewilde S, Kiger L, Burmester T, Hankeln T, Baudin-Creuza V, Aerts T, Marden MC, Caubergs R, Moens L (2001) Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J Biol Chem* **276:** 38949–38955

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Evans PR (1993) *Proceedings of the CCP4 Study Weekend, on Data Collection and Processing*, pp 114–122. Warrington, UK: CLRC Daresbury Laboratory

Farrés J, Rechsteiner MP, Herold S, Frey AD, Kallio PT (2005) Ligand binding properties of bacterial hemoglobins and flavohemoglobins. *Biochemistry* **44:** 4125–4134

Ferry JG, House CH (2006) The stepwise evolution of early life driven by energy conservation. *Mol Biol Evol* **23:** 1286–1296

Freitas TAK, Hou S, Dioum EM, Saito J, Newhouse J, Gonzalez G, Gilles-Gonzalez M-A, Alam M (2004) Ancestral hemoglobins in Archaea. *Proc Natl Acad Sci USA* **101:** 6675–6680

Galagan JE *et al* (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12:** 532–542

Giangiacomo L, Mattu M, Arcovito A, Bellenchi G, Bolognesi M, Ascenzi P, Boffi A (2001) Monomer-dimer equilibrium and oxygen binding properties of ferrous *Vitreoscilla* hemoglobin. *Biochemistry* **40:** 9311–9316

Giangiacomo L, Ilari A, Boffi A, Morea V, Chiancone E (2005) The truncated oxygen-avid hemoglobin from *Bacillus subtilis*: X-ray structure and ligand binding properties. *J Biol Chem* **280:** 9192–9202

Gibson QH (1973) The contribution of the α and β chains to the kinetics of oxygen binding to and dissociation from hemoglobin. *Proc Natl Acad Sci USA* **70:** 1–4

Hou S *et al* (2001) Globin-coupled sensors: a class of heme-containing sensors in Archaea and Bacteria. *Proc Natl Acad Sci USA* **98:** 9353–9358

Krissinel E, Henrick K (2005) Detection of protein assemblies in crystals. In *CompLife 2005*, Berthold MR *et al* (eds) LNBI 3695, pp 163–174. Berlin Heidelberg, Germany: Springer

Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. *J Mol Graph* **13:** 323–330

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK, a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291

Leslie AGM (2003) *MOSFLM User Guide, Mosflm Version 6.2.3*. Cambridge, UK: MRC Laboratory of Molecular Biology

Lessner DJ, Li L, Eejtar T, Andreev VP, Reichlen M, Hill K, Moran JJ, Karger BL, Ferry JG (2006) An unconventional pathway for reduction of CO_2 to methane in CO-grown *Methanosarcina acetivorans* revealed by proteomics. *Proc Natl Acad Sci USA* **103:** 17921–17926

Milani M *et al* (2005) Structural bases for heme binding and diatomic ligand recognition in truncated hemoglobins. *J Inorg Biochem* **99:** 97–109

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255

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

Rother M, Metcalf WW (2004) Anaerobic growth of *Methanosarcina acetivorans* C2A on carbon monoxide: an unusual way of life for a methanogenic archaeon. *Proc Natl Acad Sci USA* **101:** 16929–16934

Shelnutt JA, Song X-Z, Ma J-G, Jia S-L, Jentzen W, Medforth CJ (1998) Nonpolar porphyrins and their significance in proteins. *Chem Soc Rev* **27:** 31–41

Smerdon SJ *et al* (1995) Interactions among residues CD3, E7, E10, and E11 in myoglobins: attempts to simulate the ligand-binding properties of *Aplysia* myoglobin. *Biochemistry* **34:** 8715–8725

Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **60:** 432–438

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

Vinogradov SN, Hoogewijs D, Bailly X, Arredondo-Peter R, Gough J, Dewilde S, Moens L, Vanfleteren J (2006) A phylogenomic profile of globins. *BMC Evol Biol* **6:** 31

Vinogradov SN, Hoogewijs D, Bailly X, Mizuguchi K, Dewilde S, Moens L, Vanfleteren J (2007) A model of globin evolution. *Gene* **398:** 132–142

Wright E, Serpersu EH (2004) Isolation of aminoglycoside nucleotidyltransferase (2″)-Ia from inclusion bodies as active, monomeric enzyme. *Protein Expr Purif* **35:** 373–380

Zhang W, Phillips Jr GN (2003) Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry. *Structure* **11:** 1097–1110

Zhang W, Olson JS, Phillips Jr GN (2005) Biophysical and kinetic characterization of HemAT, an aerotaxis receptor from *Bacillus subtilis*. *Biophys J* **88:** 2801–2814