IL NUOVO CIMENTO Vol. 31 C, N. 4 Luglio-Agosto 2008
DOI 10.1393/ncc/i2008-10305-6

Crystal structure of archeal protoglobin: Novel ligand diffusion paths to the heme

A. Pesce(*)

Dipartimento di Fisica and Centro di Eccellenza per la Ricerca Biomedica Università di Genova - Via Dodecaneso, 33, 16146-Genova, Italy

(ricevuto il 29 Dicembre 2008; approvato il 13 Gennaio 2009; pubblicato online il 3 Marzo 2009)

Summary. — The protein structural adaptability of the globin fold has been highlighted by the recent discovery of 2-on-2 hemoglobins, of neuroglobin, cytoglobin, and the characterization of their three-dimensional structures. Protoglobin from *Methanosarcina acetivorans* C2A is the latest entry in the hemoglobin superfamily, adding to it new structural variability and functional complexity. The 1.3 Å crystal structure of oxygenated *M. acetivorans* protoglobin shows that, contrary to all known globins, protoglobin-specific loops and a N-terminal extension completely bury the heme within the protein matrix. Access of diatomic ligands (such as O_2, CO, and NO) to the heme is however granted by protoglobin-specific apolar tunnels that reach the heme distal site from entry sites at the B/G and B/E helix interfaces. Functionally, *M. acetivorans* dimeric protoglobin displays a selectivity ratio for O_2/CO binding to the heme that favours O_2 ligation, a property that is exceptional within the hemoglobin superfamily.

PACS 61.05.cp – X-ray diffraction.
PACS 61.90.+d – Other topics in structure of solids and liquids; crystallography.
PACS 64.70.kt – Molecular crystals.
PACS 87.15.bg – Tertiary structure.

1. – Introduction

Methanogenesis plays a pivotal role in the global carbon cycle and contributes significantly to global warming, the majority of methane in nature being produced from acetate [1]. *Methanosarcina acetivorans* C2A, a strictly anaerobic nonmotile Archaea, in contrast to most methanogens, can exploit acetate, methanol, CO_2 and CO as carbon sources for methanogenesis. Methane production in this Archaea takes place in parallel with the formation of a proton gradient that is essential for energy harvesting [2, 3].

(*) E-mail: pesce@ge.infm.it

Despite its strict anaerobic nature, *M. acetivorans* genome hosts genes that can be related to O_2 metabolism; among these, an open reading frame encodes for a "protoglobin" (NP_617780; *Ma*Pgb([1])). Pgbs are single domain heme proteins of $\sim$ 195 amino acids, related to the N-terminal domain of archaeal and bacterial globin coupled sensor proteins (GCS) [4-7]. Furthermore, sequence comparisons indicate that Pgbs, despite their 30–35% larger size, are structurally related to the single chain hemoglobins (Hbs; composed of about 150 amino acids, folded into a 3-on-3 α-helical sandwich, 12–16% residue identity to Pgbs). Although functional and evolutionary issues are openly debated [5-7], Pgbs have been proposed to facilitate O_2 detoxification *in vivo* by promoting electron transfer to O_2, or may act as CO sensor/supplier in methanogenesis. The first structural investigations on Pgb highlight large modifications of the 3-on-3 globin fold; functionally Pgbs show ligand binding properties (*vs.* O_2 and CO) that underline an almost unique behaviour within the Hb superfamily.

2. – Methods

2˙1. *Crystallization and data collection.* – *Ma*Pgb was expressed and purified as reported by Nardini *et al.* [8]. For crystallization purposes the Cys101 $\rightarrow$ Ser mutation was introduced. Crystallization of *Ma*Pgb was achieved in hanging drop vapour diffusion set-up. The protein solution at 44 mg/ml was equilibrated against a precipitant solution containing 0.4 M monobasic ammonium phosphate, 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, prior to 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, $\beta = 92.9°$. The V_M value is 2.31 Å^3 Da^{-1}, 46.9% solvent content, assuming one *Ma*Pgb molecule per asymmetric unit. All collected data were reduced and scaled using MOSFLM and SCALA [9,10], and programs from the CCP4 suite [11] (table I).

2˙2. *Structure determination and refinement.* – Multi-wavelength Anomalous Dispersion (MAD) phases, based on the heme-Fe atom anomalous signal, were determined at 1.3 Å resolution with SOLVE [12]. The electron density map was improved by solvent flattening with DM [11], yielding a figure of merit of 0.79 at 2.0 Å resolution. ARP/wARP [13] was used to extend and refine phases to 1.3 Å resolution, and for automated model building of main and side chains. The molecular model was subsequently checked manually with COOT [14] and refined to the maximum resolution using REFMAC [15]. At the end of the refinement stages (including anisotropic B-factor refinement), 1 oxygen molecule, 264 water molecules, 2 phosphate ions, and 1 glycerol molecule were located through inspection of difference Fourier maps. The final R_{factor} value is 16.0%, and R_{free} 19.3% (table I) [8].

The programs Procheck and Surfnet [16,17] were used to assess stereochemical quality and to explore protein matrix cavities. The program PISA [18] was used to identify quaternary assemblies within the crystal unit cell. Atomic coordinates and structure factors for *Ma*Pgb are available from the Protein Data Bank [19] as entries 2VEB and r2VEBsf, respectively.

([1]) Abbreviations used: protoglobin, Pgb; *Methanosarcina acetivorans* protoglobin, *Ma*Pgb; globin coupled sensor, GCS; hemoglobin, Hb; multi-wavelength anomalous dispersion, MAD; root mean square deviation, rmsd.

TABLE I. – *Data collection, phasing (MAD:* Fe*) and crystallographic refinement statistics for Ma*Pgb-O_2.

Data collection			
Space group		*C*2	
Cell dimensions:			
a, *b*, *c* (Å)		80.1, 49.3, 51.5	
α, β, γ(°)		90, 92.9, 90	
	Peak	*Inflection*	*Remote*
Wavelength (Å)	1.738	1.740	0.954
Resolution (Å)	51.4–1.7	51.4–1.7	42.0–1.3
No. of reflections	179215	81427	338891
Unique reflections	19890	18620	48188
$R_{\text{merge}}{}^{(a)}$	0.045 (0.251)	0.104(0.112)	0.048 (0.215)
$I/\sigma(I)$	35.6 (4.5)	11.6 (5.5)	24.7 (7.5)
Completeness (%)	90.3 (46.8)	84.4 (44.8)	97.8 (99.8)
Multiplicity	9.0 (5.7)	4.4 (3.1)	7.0 (7.1)
Refinement			
$R_{\text{factor}}{}^{(b)}/R_{\text{free}}{}^{(c)}$		16.0/19.0	
No. of residues/protein atoms		190 (from 6 to 195)/1603	
No. of heme groups		1	
No. of O_2 molecules		1	
No. of phosphate ions		2	
No. of glycerol molecules		1	
No. of water molecules		264	
Rmsd from ideality:			
bond lengths (Å)		0.010	
bond angles (°)		1.3	

Values in parentheses are for highest-resolution shell.

(a) *R*-merge = $\Sigma_h\Sigma_i|I_{hi} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{hi}$.

(b) *R*-factor = $\Sigma_h||F_{\text{obs}}| - |F_{\text{calc}}||/\Sigma|F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

(c) *R*-free is calculated with 10% of the diffraction data, which were not used during the refinement.

3. – Results

3˙1. *The overall structure.* – The crystal structure of oxygenated Pgb from *M. acetivorans* (*Ma*Pgb-O_2) highlights a protein fold consisting of nine main helices (labelled Z, A, B, C, E, F, G, H, and H′), partly related to those identified in Hb [20] and in GCS [21] (fig. 1A). The Z helix precedes the globin-fold conserved A-helix and is involved in protein dimerization. Indeed, the crystal packing yields a symmetric *Ma*Pgb-O_2 homodimer, whose 2086 Å^2 association interface is contributed mostly by residues belonging to the G- and H-helices (that build an intermolecular four-helix bundle), to the H′-helix, to the Z-helix, and to the BC and FG hinges (fig. 1B). Accordingly, gel filtration experiments on the expressed protein show that *Ma*Pgb elutes as a dimer in solution.

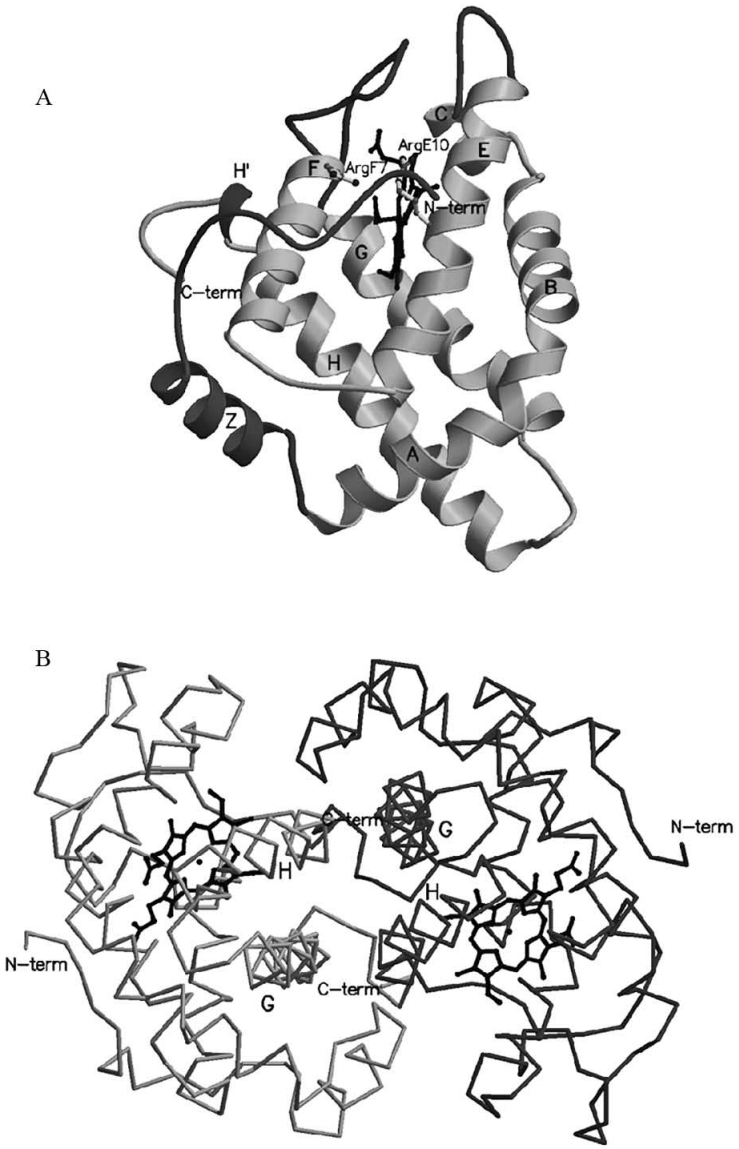

Fig. 1. – (A) The *Ma*Pgb fold. The figure highlights the secondary structure elements of the 3-on-3 Hb fold (light grey; labels A through H). The structural elements that are specific of *Ma*Pgb (relative to 3-on-3 Hbs) are displayed in dark grey. The *Ma*Pgb-specific N-terminal region, the CE and FG loops bury the heme and prevent access of small ligands to the heme distal cavity. (B) The *Ma*Pgb dimer. The two *Ma*Pgb subunits, interacting mainly through the G and H helices (helical bundle), are shown in light and dark grey stick models. All the figures are drawn with Molscript [22].

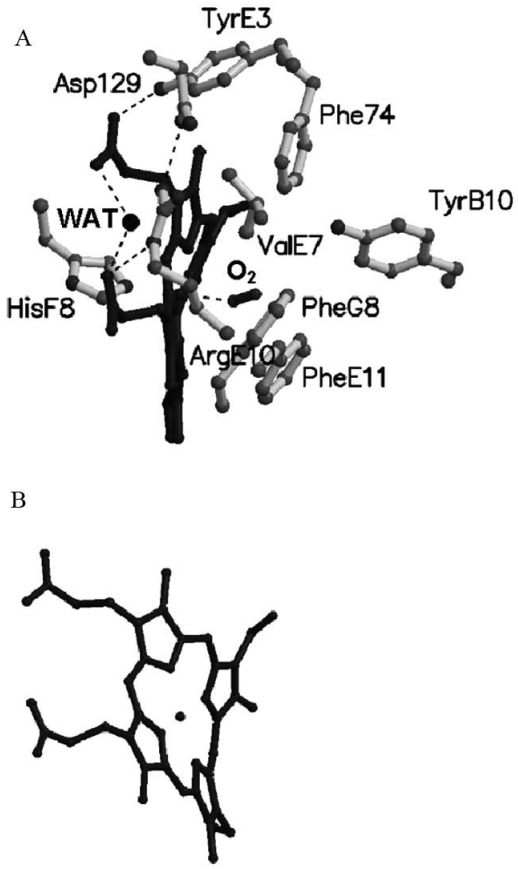

Fig. 2. – (A) *Ma*Pgb heme pocket. The figure shows the main residues building the heme distal site cavity, the O_2 molecule bound to the heme-iron atom (dark grey), and the proximal HisF8. When appropriate, the topological position of the protein residues is indicated. Hydrogen bonds are dashed; one water molecule (WAT) participating to heme propionates stabilization is shown as a black sphere. (B) Heme distortion. View of the heme highlighting its substantial planar distortion.

The Z-helix is preceded by 20 amino acids building a N-terminal loop held next to the heme propionates by hydrogen bonds that link Pro7, Gly8, Tyr9, Thr10, Ala18, and Phe20 to residues of the E and F helices in the protein α-helical core. The heme propionates are thus solvent inaccessible, and stabilized by intra-molecular 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.

3˙2. *The heme pocket*. – The heme 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 heme-Fe atom (fig. 2A). Contrary to all known Hb structures (where about 30% of the heme surface is solvent accessible), the heme 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 of the extended CE and FG loops (residues 73-81 and 122-137, respectively).

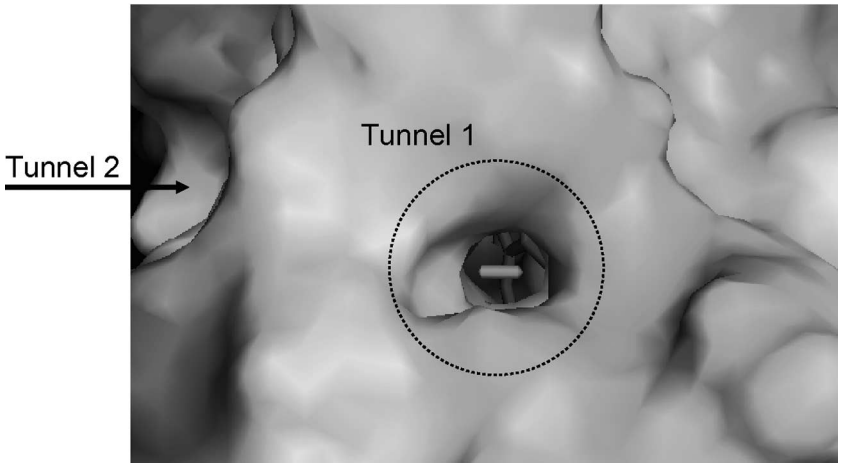

Fig. 3. – Views of "tunnel 1" and "tunnel 2" entries in *Ma*Pgb. The protein molecular surface (as defined by a 1.4 Å radius probe) is displayed. The heme group and the dioxygen molecule are displayed in sticks, as seen through the tunnel 1 aperture (highlighted in the figure with a dotted circle). The arrow shows the location of tunnel 2 entrance.

While the diatomic ligand diffusion path to the heme pocket through the E7-gate firstly reported for myoglobin [23] is thus precluded, *Ma*Pgb-O_2 displays alternate paths. A straight apolar protein matrix tunnel (about 7 Å in diameter) connecting the protein surface to the heme distal cavity is located between the B- and G-helices (tunnel 1; fig. 3). A second straight opening on the heme distal cavity is nestled between the B- and the E-helices (about 5 Å diameter), partly defined by Tyr(B10)61 (tunnel 2; fig. 3). Both tunnels host one water molecule at their solvent side aperture. Additionally, a core cavity of about 75 $Å^3$ is located between the distal and proximal heme sides; the cavity hosts four mutually hydrogen-bonded water molecules. These cavity/tunnel systems may be implicated in diatomic ligand diffusion to/from the heme, multi-ligand storage and/or (pseudo-)enzymatic reactivity.

The high experimental resolution of the *Ma*Pgb-O_2 structure provides unequivocal evidence of substantial distortion in the porphyrin ring system. While the heme-Fe atom falls slightly out of the heme plane (−0.13 Å) towards the proximal site, out-of-plane deviations of ±0.5–0.6 Å affect the four heme pyrrole rings (fig. 2B). Specific *Ma*Pgb/heme contacts include residue-pyrrole π-π interactions, hydrogen bonds and salt bridges at the propionates, and interaction with a buried water molecule. Since deviation of porphyrins from planarity is energetically unfavourable, the marked heme distortion suggests a role of *Ma*Pgb in modulating heme reactivity and axial ligand affinity, thus providing a possible mechanism for differentiating between CO and O_2 binding, the overall O_2 affinity for *Ma*Pgb ($P_{50} \sim 2.5$ mmHg) being higher than that of CO ($P_{50} \sim 10$ mmHg), a ligand binding behaviour that is exceptional within the Hb superfamily [8]. This mechanism seems to be particularly important in Pgb since in the *Ma*Pgb-O_2 structure the heme-bound O_2 molecule is not stabilized by any direct H-bonds to the protein (fig. 2A).

4. – Discussion

From the structural and functional viewpoint Pgbs appear to be one more successful engineering experiment within the Hb superfamily, which already includes single domain (3-on-3) Hbs, flavohemoglobins, truncated (2-on-2) Hbs, and GCSs [7]. Structural modulation of the 3-on-3 fold in Pgb translates into novel access routes to the heme, into unique modulation of heme structure/reactivity, and into a specific quaternary assembly. Functionally such structural properties code for a strikingly modified O_2/CO selectivity ratio [8]. Given the unique 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 [7], the above results support the idea that Pgbs are the evolutionary archetype of GCSs, but not of all globins.

* * *

I would like to thank all the people who contributed to different stages of this work: M. Nardini and M. Bolognesi (University of Milan, Italy), L. Thijs, S. Dewilde and L. Moens (University of Antwerp, Belgium), P. Ascenzi (University of "Roma Tre", Italy) and M. Coletta (University "Tor Vergata", Rome, Italy).

REFERENCES

[1] Galagan J. E. *et al.*, *Genome Res.*, **12** (2002) 532.
[2] Rother M. and Metcalf W. W., *Proc. Natl. Acad. Sci., USA*, **101** (2004) 16929.
[3] Lessner D. J., Li L., Eejtar T., Andreev V. P., Reichlen M., Hill K., Moran J. J., Karger B. L. and Ferry J. G., *Proc. Natl. Acad. Sci. USA*, **103** (2006) 17921.
[4] Hou S., Freitas T. A. K., Larsen R. W., Piatibratov M., Sivozhelezov V., Yamamoto A., Meleshkevitch E. A., Zimmer M., Ordal G. W. and Alam M., *Proc. Natl. Acad. Sci. USA*, **98** (2001) 9353.
[5] Freitas T. A. K., Hou S., Dioum E. M., Saito J., Newhouse J., Gonzalez G., Gilles-Gonzalez M.-A. and Alam M., *Proc. Natl. Acad. Sci. USA*, **101** (2004) 6675.
[6] Vinogradov S. N., Hoogewijs D., Bailly X., Arredondo-Peter R., Gough J., Dewilde S., Moens L. and Vanfleteren J., *BMC Evol. Biol.*, **6** (2006) 31.
[7] Vinogradov S. N., Hoogewijs D., Bailly X., Mizuguchi K., Dewilde S., Moens L. and Vanfleteren J., *Gene*, **398** (2007) 132.
[8] Nardini M., Pesce A., Thijs L., Saito J. A., Dewilde S., Alam M., Ascenzi P., Coletta M., Ciaccio C., Moens L. and Bolognesi M., *EMBO Rep.*, **9** (2008) 157.
[9] Leslie A. G. M., in *MOSFLM User Guide, Mosflm Version 6.2.3*, MRC Laboratory of Molecular Biology, Cambridge, UK (2003).
[10] Evans P. R., in *Proceedings of the CCP4 Study Weekend, on Data Collection and Processing*, CLRC Daresbury Laboratory, UK (1993), pp. 114-122.
[11] Collaborative Computational Project Number 4, *Acta Crystallogr. D Biol. Crystallogr.*, **50** (1994) 760.
[12] Terwilliger T. C. and Berendzen J., *Acta Crystallogr. D Biol. Crystallogr.*, **55** (1999) 849.
[13] Perrakis A., Morris R. and Lamzin V., *Nature Struct. Biol.*, **6** (1999) 458.
[14] Emsley P. and Cowtan K., *Acta Crystallogr. D Biol. Crystallogr.*, **60** (2004) 2126.
[15] Murshudov G. N., Vagin A. A. and Dodson E. J., *Acta Crystallogr. D Biol. Crystallogr.*, **53** (1997) 240.
[16] Laskowski R. A., Macarthur M. W., Moss D. S. and Thornton J. M., *J. Appl. Crystallogr.*, **26** (1993) 283.
[17] Laskowski R. A., *J. Mol. Graph.*, **13** (1995) 323.

[18] Krissinel E. and Henrick K., in *Detection of Protein Assemblies in Crystals*, in Berthold M. R. *et al.* (Editors), *CompLife 2005*, LNBI 3695 (Springer-Verlag Berlin Heidelberg) pp. 163-174.
[19] Berman H. M., Westbrook J., Fenz Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N. and Bourne P. E., *Nucleic Acids Res.*, **28** (2000) 235.
[20] Bolognesi M., Bordo D., Rizzi M., Tarricone C. and Ascenzi P., *Prog. Biophys. Molec. Biol.*, **68** (1997) 29.
[21] Zhang W. and Phillips G. N. jr., *Structure*, **11** (2003) 1097.
[22] Kraulis P. J., *J. Appl. Crystallogr.*, **24** (1991) 946.
[23] Bolognesi M., Cannillo E., Ascenzi P., Giacometti G. M., Merli A. and Brunori M., *J. Mol. Biol.*, **158** (1982) 305.