BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

The structure of the giant haemoglobin from *Glossoscolex paulistus*

José Fernando Ruggiero Bachega,[a] Fernando Vasconcelos Maluf,[a] Babak Andi,[b] Humberto D'Muniz Pereira,[a] Marcelo Falsarella Carazzollea,[c,d] Allen M. Orville,[b,e] Marcel Tabak,[f] José Brandão-Neto,[g] Richard Charles Garratt[a] and Eduardo Horjales Reboredo[a]*

[a]Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil, [b]Photon Sciences Directorate, Brookhaven National Laboratory, Upton, New York, USA, [c]Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil, [d]Centro Nacional de Processamento de Alto Desempenho, Universidade Estadual de Campinas, Campinas, Brazil, [e]Biosciences Department, Brookhaven National Laboratory, Upton, New York, USA, [f]Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, Brazil, and [g]Diamond Light Source, Harwell, England. *Correspondence e-mail: horjales@ifsc.usp.br

Received 12 November 2014
Accepted 16 March 2015

Edited by K. Miki, Kyoto University, Japan

Keywords: erythrocruorins; *Glossoscolex paulistus*; giant extracellular haemoglobin; hexagonal bilayer.

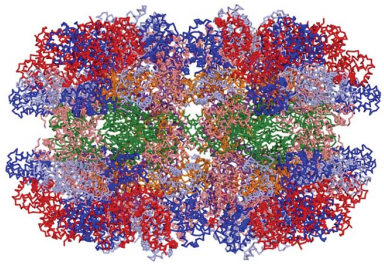
PDB references: giant haemoglobin from *G. paulistus*, 4u8u; isolated *d* chain, 4wch

Supporting information: this article has supporting information at journals.iucr.org/d

The sequences of all seven polypeptide chains from the giant haemoglobin of the free-living earthworm *Glossoscolex paulistus* (HbGp) are reported together with the three-dimensional structure of the 3.6 MDa complex which they form. The refinement of the full particle, which has been solved at 3.2 Å resolution, the highest resolution reported to date for a hexagonal bilayer haemoglobin composed of 12 protomers, is reported. This has allowed a more detailed description of the contacts between subunits which are essential for particle stability. Interpretation of features in the electron-density maps suggests the presence of metal-binding sites (probably Zn^{2+} and Ca^{2+}) and glycosylation sites, some of which have not been reported previously. The former appear to be important for the integrity of the particle. The crystal structure of the isolated *d* chain (*d*-HbGp) at 2.1 Å resolution shows different interchain contacts between *d* monomers compared with those observed in the full particle. Instead of forming trimers, as seen in the complex, the isolated *d* chains associate to form dimers across a crystallographic twofold axis. These observations eliminate the possibility that trimers form spontaneously in solution as intermediates during the formation of the dodecameric globin cap and contribute to understanding of the possible ways in which the particle self-assembles.

1. Introduction

Haemoglobins (Hbs) are present in organisms from all five kingdoms of life and have been widely studied, allowing detailed descriptions of their functional properties. Invertebrates have a wide variety of haemoglobin types ranging from monomeric structures to megadalton multimeric complexes. These molecules occur in different anatomical sites with some being extracellular, a fact which usually confers them with special characteristics when compared with those which are present within an intracellular environment (Weber & Vinogradov, 2001). A strong motivation to study these extracellular haemoglobins, especially the giant multimeric complexes such as erythrocruorins, is related to their potential application in medicine as blood substitutes. Studies performed on the giant Hbs from *Lumbricus terrestris* (HbLt; Hirsch *et al.*, 1997; Elmer & Palmer, 2012) and *Arenicola marina* (HbAm; Rousselot *et al.*, 2006; Harnois *et al.*, 2009) have demonstrated their greater resistance to oxidation and increased stability compared with human haemoglobin, highlighting their potential for medical applications (Elmer & Palmer, 2012).



The giant extracellular Hbs of annelids, also known as hexagonal bilayer haemoglobins (HBL Hbs) or erythrocruorins, together with chlorocruorins (a variant of HBL Hbs restricted to four families of polychaetes), are recognized to form a single group according to their physicochemical properties (Vinogradov, 1985). Early studies using electron microscopy applied to haemoglobins from annelids revealed that they are huge complexes containing a large number of polypeptide chains divided into two different types, globins and linkers, which together form a characteristic quaternary structure. This consists of two superimposed hexagonal layers displaying an overall D_6 (or 622) symmetry (Kapp *et al.*, 1982). The structural characterization of the Hb of *L. terrestris* (HbLt) by electron microscopy revealed that the hexagonal structure is comprised of multimeric units called protomers which present a pseudo-threefold axis of symmetry. 12 such protomers are necessary to complete the hexagonal bilayer, and the haemoglobin from *L. terrestris* presents the two layers staggered with respect to one another by approximately 16°. This arrangement is called the type I (or non-eclipsed) form. A similar quaternary structure has been found in giant haemoglobins from marine organisms, such as *Riftia pachyptila* (a vestimentiferan), *Alvinella pompejana* (a polychaete) and the clorocruorin from *Eudistylia vancouverii* (also a polychaete) (De Haas, Biosset *et al.*, 1996; De Haas, Taveau *et al.*, 1996; De Haas, Zal *et al.*, 1996). However, the structures from *A. pompejana* (HbAp) and *Arenicola marina* (HbAm) (Royer *et al.*, 2007) lack the 16° stagger, instead forming the so-called eclipsed or type II arrangement.

Vestimentiferans have two types of extracellular Hbs. One, named 'heavy', has a molecular mass of between 3.0 and 4.0 MDa and corresponds to a hexagonal bilayer structure, while the other, named 'light', is of the order of 400 kDa. Pogonophorans (a group of animals close to vestimentiferans and polychaetes), on the other hand, present only the 'light' form. Two crystal structures have been obtained for these lighter complexes: those from *Oligobrachia mashikoi*, Hb_LOm (Numoto *et al.*, 2005, 2008), and *R. pachyptila* (Flores *et al.*, 2005). Each is composed of six copies of four globin subunits (A1, A2, B1 and B2) forming a pair of dome-shaped structures associated to form a hollow spherical assembly. The oxygen-binding properties of these haemoglobins are qualitatively similar to those of annelid giant Hbs.

The crystal structures of the erythrocruorins from *L. terrestris* (HbLt) at 3.5 Å resolution (Royer *et al.*, 2006) and *A. marina* (HbAm) at 6.2 Å resolution (Royer *et al.*, 2007) have been reported and are representatives of the type I and type II arrangements, respectively. With the crystal structure of HbLt, Royer and coworkers greatly contributed to clarifying the relative stoichiometry of the constituent subunits and also the hierarchical packing of the complex. In the structure there are four different types of haem-containing globin chains, named *a*, *b*, *c* and *d*, which together form a heterotetramer in which the *a*, *b* and *c* chains are linked by disulfide bonds. The heterotetramer is then repeated three times to form a structure with *C*3 symmetry called the dodecamer or 'cap', with $(abc)_3(d)_3$ stoichiometry. This dome-shaped structure, with a molecular mass of approximately 200 kDa, has also been characterized in isolation by X-ray crystallography. The structure solved at 2.4 Å resolution (Strand *et al.*, 2004) gave a detailed model of the cap, showing that it is very similar to those found in the extracellular 'light' Hbs of vestimentiferans and pogonophorans.

In the case of HbLt, as well as other hexagonal bilayer haemoglobins and chlorocruorins, the dodecameric cap is associated with a heterotrimer of linker chains formed of a single copy each of L1, L2 and L3. This structure, with stoichiometry $[(abc)_3(d)_3]$(L1L2L3), forms the previously mentioned protomer. The globular parts of the linkers, together with the cap, form the head of the protomer, from which the tails of the linkers protrude into the centre of the particle, generating a mushroom-like appearance. In summary, therefore, the hexagonal bilayer is composed of 12 protomers, including a total of 144 globin chains and 36 linkers. The interactions formed by the latter are primarily responsible for stabilizing the structure of the entire particle.

The oxygen-binding properties of erythrocruorins exhibit features that differ significantly from those of vertebrate tetrameric haemoglobins. Whilst the cooperative oxygen-binding properties of the latter are enhanced by small organic anions or chloride, in the case of giant haemoglobins the best characterized heterotropic effectors are inorganic cations. In general, divalent cations are more effective than monovalent species according to the following series: $Ba^{2+} > Ca^{2+} > Sr^{2+} > Mg^{2+} > Li^{+} > Na^{+} > K^{+}$ (Weber & Vinogradov, 2001). Although the crystal structure of HbLt provided valuable information concerning the presence of calcium at specific binding sites (Royer *et al.*, 2007), little is known about the effects of other metal ions such as Zn^{2+} and Cu^{2+}, which have been identified in samples from many oligochaetes, including HbLt (Standley *et al.*, 1988).

The giant extracellular haemoglobin from *Glossoscolex paulistus* (HbGp), an earthworm found in southeast Brazil, is similar in structure to HbLt and is one of the best characterized erythrocruorins (Santiago, Carvalho, Domingues, Carvalho *et al.*, 2010; Santiago, Carvalho, Domingues, Santos *et al.*, 2010; Carvalho *et al.*, 2009, 2013). A preliminary molecular-replacement solution of the HbGp complex, using a protomer from HbLt as a search model, has been reported previously (Bachega *et al.*, 2011) and shown to present the staggered type I arrangement of protomers. Nevertheless, until the present work, the only sequence available for HbGp was that of the globin *d* subunit (Bosch Cabral *et al.*, 2002), making refinement and interpretation of the crystal structure impossible. Here, we report the amino-acid sequences of all seven chains which comprise the complex, together with its structure refined at 3.2 Å resolution.

2. Materials and methods

2.1. Sequencing of the HbGp subunits

Total RNA (<10 µg) was extracted from *G. paulistus* using the RNeasy kit from Qiagen. The quality and quantity of RNA

was determined by monitoring the absorbance at 260 nm and on 1% agarose gels under denaturing conditions. It was subsequently subjected to cDNA synthesis prior to sequencing. The cDNA was sequenced using a high-throughput sequencing platform (Illumina HiSeq 2000), generating 2 × 100 bp paired-end reads (250 bp insert length). Sequencing was performed at the High Throughput Sequencing Facility (HTSF) of the University of North Carolina, Chapel Hill, USA. *De novo* assembly of paired-end reads was performed using the Trinity assembler (Grabherr *et al.*, 2011), setting the maximum insert parameter to 300 bp and considering a minimum contig size of 500 bp. The polypeptide chains of *L. terrestris* haemoglobin were used to identify homologous transcripts within the *G. paulistus* transcriptome assembly using *BLASTx* (Altschul *et al.*, 1997), applying an *E*-value cutoff of 10^{-5}. The hits were subsequently manually verified.

2.2. Structure determination and crystallographic refinement of the full biological particle

Some of the details of the data collection and structure solution have been described in our preliminary report (Bachega *et al.*, 2011). Briefly, the whole HbGp complex was purified directly from adult earthworms and crystallized as described previously (Bachega *et al.*, 2011). X-ray diffraction images were collected at 93 K with an ADSC Quantum 315 detector using synchrotron radiation of wavelength 1.000 Å on beamline X29C of the NSLS (Brookhaven National Laboratory, Upton, New York, USA).

Diffraction data processing to 3.2 Å resolution was performed using *iMosflm* (Leslie, 2006; Battye *et al.*, 2011) and *SCALA* from the *CCP*4 package (Winn *et al.*, 2011). The data-collection and processing statistics have been published previously (Bachega *et al.*, 2011). Molecular replacement with *Phaser* (McCoy *et al.*, 2007) employing a single protomer from HbLt (PDB entry 2gtl; Royer *et al.*, 2006; including both main and side chains) was used to determine the locations of the three protomers within the asymmetric unit. The haem groups were omitted during this procedure. After amino-acid substitution (where necessary), *PHENIX* (Adams *et al.*, 2002) was used to refine the structure. Three runs of three cycles each were alternated with local real-space refinement and visual evaluation with *Coot* (Emsley & Cowtan, 2004). It should be borne in mind that the asymmetric unit contains three protomers, each of which is composed of three (*abcd*) tetramers and an (L1L2L3) trimer of linkers. The large size of the structure and the limited resolution (3.2 Å) therefore required some special care as follows.

(i) Noncrystallographic symmetry (NCS) was essential to inhibit over-refinement. During the first three cycles using *PHENIX* we tested two different possible sets of NCS. The first used a full protomer (three globin tetramers and three linkers) as the reference structure, whilst the other used only one tetramer and three linkers as reference structures (the seven unique chains). The first set of NCS did not generate significant differences among the three equivalent globins within the protomer (for example among the different copies of the *a* chain) and the refinement evolution was slower than the second set. Thus, we continued the refinement using NCS based on one sequence/one reference structure. This procedure required 13 Gb of RAM memory to run. We used a computer with an Intel CPU Core I7 2600 (Sandy Bridge) with 16 Gb RAM memory running Linux Ubuntu. Under these conditions the X-ray/stereochemistry weight was adjusted to simultaneously optimize the geometry of the model and the fit to the X-ray diffraction data.

(ii) During the first *PHENIX* run, a larger radius was defined in the bulk-solvent mask calculation in order to avoid losing surface details such as the sugars that clearly became well defined later in the refinement process.

(iii) During the final steps of refinement we reduced the NCS restraints from medium to loose without changes to the observed r.m.s.d.s.

2.3. Crystallization, structure determination and crystallographic refinement of the isolated *d* chain

The isolated *d* chain of HbGp (*d*-HbGp) was purified according to a previously described method (Cabral *et al.*, 2002). Oxy-state crystals were obtained by hanging-drop vapour diffusion at 291 K in the presence of 1.4 *M* sodium citrate, 50 m*M* Tris–HCl pH 7.5. The addition of 10% ethylene glycol to the original growth conditions was used for cryoprotection. Diffraction data from crystals of *d*-HbGp were collected on beamline I04-1 at Diamond Light Source equipped with a Pilatus 2M detector using a wavelength of 0.9200 Å, an oscillation range of 0.4° and a 1 s exposure time per image at a temperature of 80 K. The search model used for molecular replacement with *Phaser* was the isolated *d* subunit from the full particle structure solved at 3.2 Å resolution and was refined using *PHENIX*, reserving 10% of the reflections for the calculation of $R_{\rm free}$.

3. Results and discussion

3.1. Sequence determination and analysis

The *ab initio* transcriptome assembly generated 97 639 contigs larger than 500 bp representing the most expressed isoforms of *G. paulistus*. The *L. terrestris* haemoglobin was used to identify homologous transcripts within *G. paulistus*, resulting in six complete globin chains (Fig. 1) and five complete linker chains (Fig. 2), all of which included N-terminal sequences rich in hydrophobic residues (Supplementary Table S1). These were predicted to be signal peptides by the *SignalP* 4.1 server (http://www.cbs.dtu.dk/services/SignalP/). The different chains were assigned according to their sequence identity when compared with their counterparts in HbLt. The six globin sequences correspond to subunits *a*1, *a*2, *b*, *c*, *d*1 and *d*2, where *a*1 and *a*2 are isoforms of the *a* subunit and *d*1 and *d*2 are isoforms of the *d* subunit. HbLt presents isoforms of the *d* chain (Xie *et al.*, 1997) which share 80% sequence identity, whereas in *G. paulistus* this value falls to only 58%. The isoform we name *d*1 is equivalent

to the *d* subunit sequence reported previously (Bosch Cabral *et al.*, 2002). As expected, the half-cystine residues which form both intramolecular and intermolecular disulfide bridges are conserved in all isoforms.

The complete data set of RNA-seq reads has been deposited in the SRA with accession No. SRR1519963.

Overall, the globin chains share sequence identities of at least 45%, similar to that observed for HbLt. However, comparisons between equivalent globin subunits in HbGp and HbLt show a greater level of sequence identity for the *a*, *b* and *c* chains than was previously estimated based on the *d*1 sequence reported by Cabral and coworkers. The residues around the sixth coordination site (37, 51, 71 and 113) form a hydrophobic pocket which is highly conserved in all chains and is also present in human globins (Fig. 1).

Of the sequences identified as linkers, three were readily classified as L1, L2 and L3. These sequences share about 60% identity with their homologues in HbLt but only approximately 30% amongst themselves. A fourth sequence was also identified which is similar to that of L4 in HbLt. Here, we consider it to be an isoform of the L3 subunit and have named it L3b. The sequence identity between L3 and L3b is 43%, and neither share more than 30% sequence identity with the remaining linkers. The fifth and final sequence identified

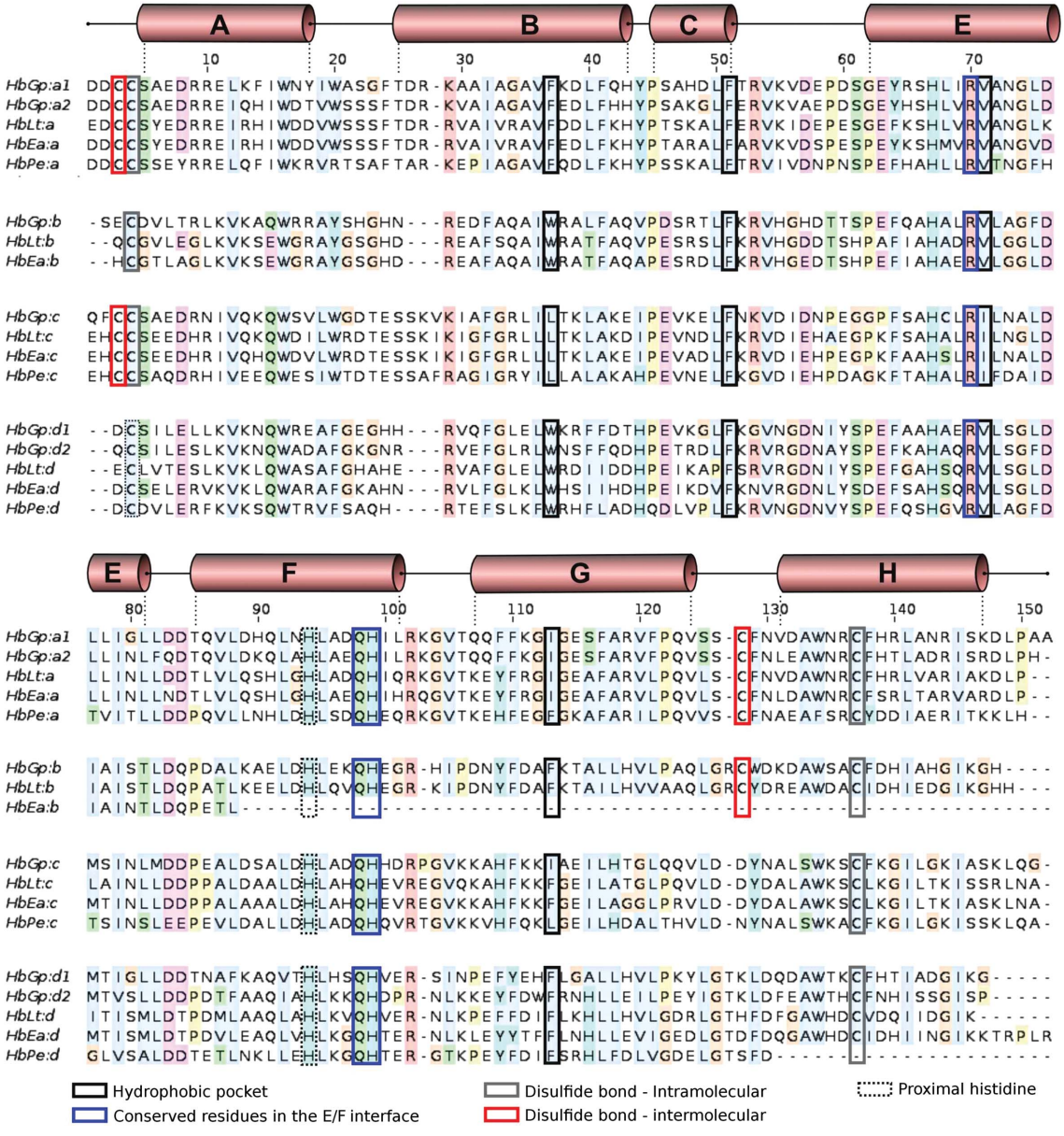


Figure 1
Sequence alignment of the HbGp globin chains. The six different globin sequences from *G. paulistus* identified in this study (*a*1, *a*2, *b*, *c*, *d*1 and *d*2) have been aligned with their homologues from *L. terrestris* (HbLt), *Eisennia andrei* (HbEa) and *Perionyx excavatus* (HbPe) according to the four different groups. The *G. paulistus* haemoglobin shows the presence of two isoforms for the *a* and the *d* chains. Standard nomenclature is used to identify the α-helices (A–H) as cylinders. Functionally important residues are boxed and their significance is indicated on the figure. The background colours code for well conserved residues: basic (red), acidic (pink), hydrophilic (green), glycine (orange), proline (yellow) and hydrophobic (blue).

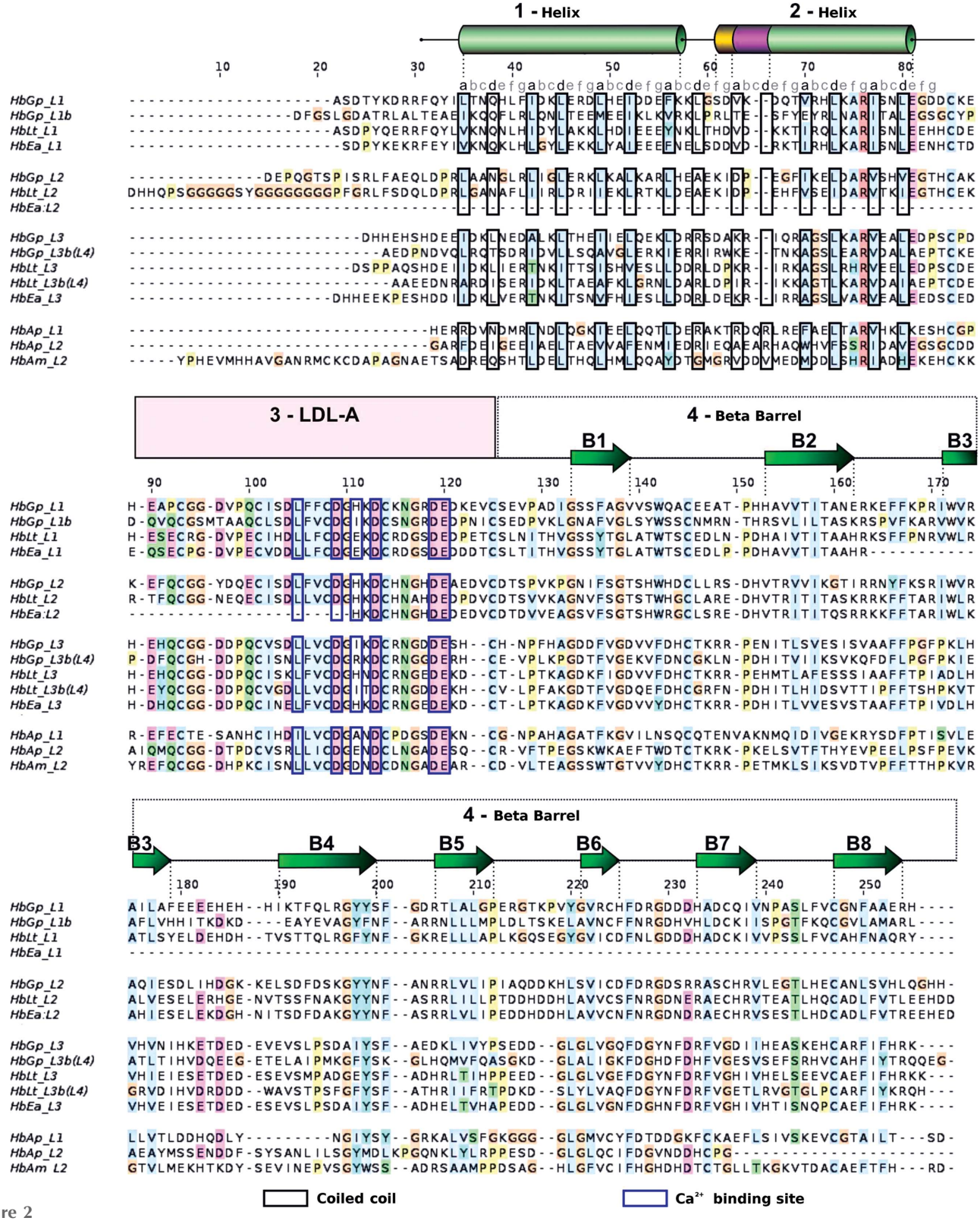


Figure 2
Sequence alignment of the HbGp linker chains. The five linker sequences identified in this study are shown aligned with homologues from *L. terrestris* (HbLt) and *E. andrei* (HbEa) as well as the marine species *A. pompejana* (HbAp) and *A. marina* (HbAm). Owing to the presence of isoforms. two distinct sequences were identified for the first and third linkers. Those which were most similar to the HbLt sequences have been named L1 and L3, and the remaining isoforms have been named L1b and L3b, respectively. Secondary-structure elements are indicated as cylinders and arrows. The structure is divided into four domains. Domains 1 and 2 correspond to individual helices and form the coiled-coil stalk of the linker trimer. Domains 3 and 4 are the LDL-A domain and the *β*-barrel domain, respectively. The phase of the coiled coil is indicated by the letters above the sequence and the hydrophobic residues at positions *a* and *d* are boxed, as are residues of the LDL-A domain which are involved in Ca^{2+} binding. The background colour code is as in Fig. 1.

shares 35 and 34% sequence identity with L1 and L2, respectively, and is thus ambiguous with respect to its most appropriate classification. Here, it has arbitrarily been named L1b.

A comparison of the predicted physicochemical properties of the sequences of HbGp and HbLt shows that there are no major differences with respect to the total masses and isoelectric points (Supplementary Table S1). The N-terminal regions of the three linker subunits are characterized by sequences which present heptad repeats typical of coiled coils (*abcdefg*), with hydrophobic residues at *a* and *d*. Fig. 2 shows the alignment of the HbGp linker sequences compared with those from other annelids, highlighting the residues that form the coiled coil. Type I (uneclipsed) erythrocruorins are characterized by a deletion at positions 65 and 66 which results in a breakdown in the phase of the heptad repeats, whereas type II erythrocruorins (HbAp and HbAm) retain the phase of the coiled coil throughout. This is consistent with low-resolution maps of the structure of HbAm (Royer *et al.*, 2007), which show a contiguous α-helix rather than one divided into two segments as seen in the type I structures (see below). The presence of gaps at positions 65 and 66 in all HbGp linker sequences is thus consistent with it being a type I erythrocruorin, as confirmed by the crystal structure described below. The remaining two domains of the linkers (the LDL-A domain and the β-barrel domain) can be readily delimited from the alignment with the HbLt sequences. A comparison between the experimental masses obtained by MALDI-TOF MS analysis (Oliveira *et al.*, 2007; Carvalho *et al.*, 2011) and those derived from the sequences reported here is given in the Supporting Information.

3.2. Crystallographic refinement

A preliminary report has already described the basic structure solution (Bachega *et al.*, 2011) and a brief summary is given here in the Supporting Information. The initial electron-density maps ($2F_o - F_c$) showed clear signs of many side chains and haem groups (which had been removed from the search model), enhancing confidence in the MR solution. The asymmetric unit (three protomers) consists of a total of 45 subunits (36 globins and nine linkers) totalling 7200 amino-acid residues. In the *a*, L1, L2 and L3 subunits, the first two, three, 17 and five N-terminal residues, respectively, are not observed in the electron-density maps. In all other subunits the main chain is complete if we use the predicted cleavage position of the signal peptide as a criterion for determining the N-terminus (Supplementary Table S1). Where different isoforms existed, the choice of sequence was taken to be that which was most consistent with the calculated electron density. Thus, all *a* and *d* globin subunits were modelled using sequences *a*2 and *d*1, respectively (Supplementary Fig. S1). Linkers L1 and L3 were best accounted for by the sequences labelled as such in Fig. 2, with the L1b and L3b isoforms considered to be less compatible with the maps.

The limited resolution and the large number of atoms in the asymmetric unit (59 091) demanded the use of NCS restraints and group temperature-factor refinement. The final R and R_{free} values were 21.7 and 23.5%, respectively. The weights applied in order to maintain standard stereochemistry during refinement can be considered satisfactory, since R_{free} fell at each stage of the refinement process, reaching an acceptable final value (Kleywegt & Jones, 2002), and simultaneously resulted in good r.m.s.d.s for both bond lengths and angles. The φ/ψ distribution is also consistent with that expected for structures of similar resolution. The main data for the refinement process are presented in Table 1.

In the region of the haem groups, the electron density is sufficiently clear to determine the orientation of the haem plane in all subunits. Since the sample crystallized was the cyano form of HbGp, a CN moiety was modelled as the sixth ligand of the Fe atoms. However, it should be mentioned that this density is insufficiently clear to be able to unambiguously establish the precise chemical nature of the sixth ligand. This is owing to a series of factors, including the proximity of the density to the iron, the poor resolution of the data and redox effects owing to the ionizing radiation (Andi *et al.*, in preparation). The electron-density map also confirmed the presence of Ca^{2+} ions ($2F_o - F_c$ density peak intensities with

Table 1
Details of the crystallographic refinement.

Standard crystallographic parameters are given for both the full particle and the isolated *d* chain. Values in parentheses are for the outer resolution shell.

	Hb	Monomer
Data collection		
Space group	*I*222	*I*222
Unit-cell parameters (Å)	*a* = 272.68, *b* = 319.90, *c* = 333.18	*a* = 53.11, *b* = 63.15, *c* = 78.39
Detector	ADSC Quantum 315	Pilatus 2M
X-ray source	X29C, NSLS	I04-1, DLS
Wavelength (Å)	1.00	0.9200
Resolution range (Å)	49.65–3.20 (3.37–3.20)	43.97–2.05 (2.11–2.05)
Multiplicity	4.7 (4.2)	4.3 (4.3)
R_{merge} (%)	11.2 (50.0)	6.8 (97.7)
Completeness (%)	99.8 (99.2)	99.5 (99.3)
Total reflections	1105247 (145116)	36867 (2836)
Unique reflections	237062 (34189)	8521 (657)
$\langle I/\sigma(I)\rangle$	11.2 (3.1)	19.5 (2.5)
Refinement parameters		
Reflections used in refinement	236866	8511
R (%)	21.65	20.96
R_{free} (%)	23.52	23.39
No. of protein atoms	57081	1182
No. of ligand atoms	2082	54
B factor from Wilson plot (Å^2)	60.4	27.60
Cruickshank DPI (Å)	0.21	0.28
Coordinate error (maximum-likelihood)	0.41	0.25
Phase error	24.12	25.21
Ramachandran plot		
Favoured (%)	95.58	97.83
Allowed (%)	3.40	2.17
Outliers (%)	1.01	0.0
All-atom clashscore	12.78	10.82
R.m.s.d. from ideal geometry		
Bond lengths (Å)	0.012	0.005
Bond angles (°)	0.962	0.967
PDB code	4u8u	4wch

values of 6.5 ± 0.5σ for all nine peaks in the asymmetric unit) coordinated to the LDL-A domain in all linker chains (Fig. 3*a*). These are present at equivalent positions to those found in the linkers of HbLt. Electron density for all three L3-type subunits also revealed two further peaks indicating the presence of other heavy atoms (Figs. 3*b* and 3*c*), one exclusive to L3 and a second at the interface between L3 and L2. These were modelled as Zn^{2+} ions after taking into account the nature of the residues that comprise their chemical environment (particularly the first coordination sphere) and owing to the fact that Zn^{2+} was detected in the sample by inductively coupled plasma atomic emission analysis (data not shown). The $2F_o - F_c$ density peak intensities presented values of 6.5 ± 0.3σ for all three peaks in the interior of L3 and 9.2 ± 0.2σ for the three peaks at the L3–L2 interface. This difference might indicate an occupancy of less than 1.0 for the internal L3 site. By comparison, all Fe atoms from the different haems present a $2F_o - F_c$ density of higher than 9.5σ. On the other hand, the only Zn^{2+} ion present in the crystal structure of HbLt lies in the cavity ('mouth') formed by the β-barrel of subunit L2 (see below). The amino-acid residues which compose this site are conserved in the L2 sequence of HbGp, and the $F_o - F_c$ difference maps indicate the presence of a scattering ion at this position. However, the chemical environment is rich in positively charged residues (Arg123 and Arg209 in *L. terrestris* and Arg216 and Arg130 in *G. paulistus*) and therefore seems inconsistent with the presence of a Zn^{2+} cation. This density was left uninterpreted, but we believe that it is more likely to be an anion (such as chloride or phosphate) than a metal, as reported previously.

The sequences of L2 and L3 are predicted to be *N*-glycosylated based on the amino-acid sequences. However, the electron density is only able to confirm the glycosylation of L3 subunits at Asn121, which was modelled as an *N*-acetylglucosamine (Fig. 3*d*). The glycosylation and the presence of the Zn^{2+} site are strong evidence that it is the L3 sequence which is dominant in the crystallographic model of HbGp, since the sequence of the L3b isoform lacks both of these features. The carbohydrate moiety points inwards towards a cavity lined by the heads of four protomers, two from each layer (see Supplementary Fig. S2). This space arises where the coiled-coil spokes of the linker chains leave the head and are directed towards the centre of the structure, creating a continuous doughnut-shaped cavity which runs around the particle. The extent to which this cavity may be occupied will depend on the size of the carbohydrate moieties, which is unknown at present.

3.3. Complex organization (structural overview)

Like HbLt, the structure of HbGp is a type I non-eclipsed bilayer (Royer *et al.*, 2000) in which each layer consists of six protomers. This is consistent with the idea that all earthworms present erythrocruorins of this type (Weber & Vinogradov, 2001). The biological unit (or full particle) is a 180-subunit structure (144 globins and 36 linkers) with an overall diameter of approximately 290 Å and a height of 190 Å (Fig. 4). The protomers are oligomeric structures with a mushroom-like format which is approximately 130 Å in length. For descriptive purposes, the protomer can be conveniently divided into two parts: the cap (a dodecamer of globins) and a heterotrimer of linkers. The cap is a dome-shaped structure composed of a trimer of tetramers, $(abcd)_3$, where subunits a_1, b_1 and c_2 (*c* from a neighbouring tetramer; see below) are connected by disulfide bonds. The trimer of linkers is a stoichiometric

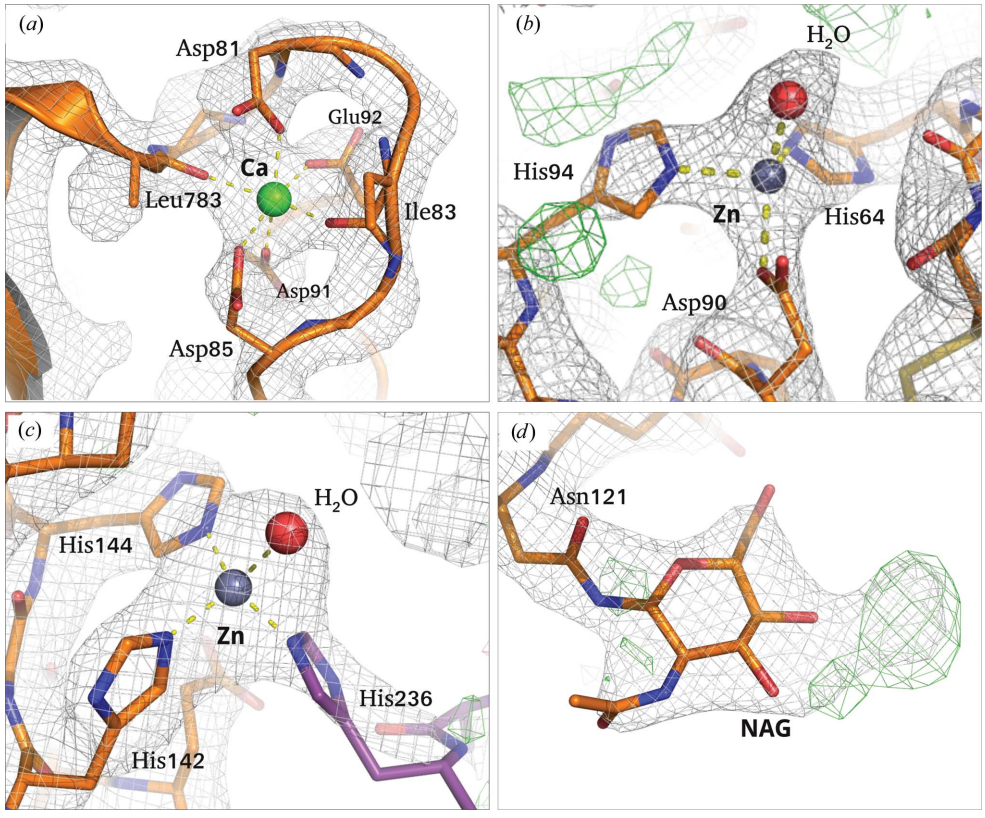


Figure 3
Selected regions of electron density. Grey densities represent $2F_o - F_c$ maps (contour level 1σ) and green densities represent positive $F_o - F_c$ maps (contour level 3σ). (*a*) The Ca^{2+}-binding site within the LDL-A domain observed in all linker chains, exemplified here by L3, (*b*) the L3 Zn^{2+}-binding site, (*c*) a second Zn^{2+} site at the interface between L3 and L2 and (*d*) the glycosylation site on Asn121 of L3.

association of L1, L2 and L3 subunits (labelled *M*, *N* and *O*, respectively, in the PDB file), forming a funnel-shaped structure whose 'head' (the C-terminal portion, composed of an LDL-A domain and a β-barrel domain) interacts with the dodecameric globin cap. The N-terminal tails of the trimer form coiled coils which point, like spokes, towards the centre of the structure. Fig. 4 establishes the standard colours used in this work to distinguish between the different kinds of subunit as well as the various levels within the structural hierarchy which will be described below.

3.4. Globins and cap formation

In HbGp the globin chains present the typical fold originally described by Perutz (see, for example, Perutz, 1997). The F helix, however, is significantly larger when compared with

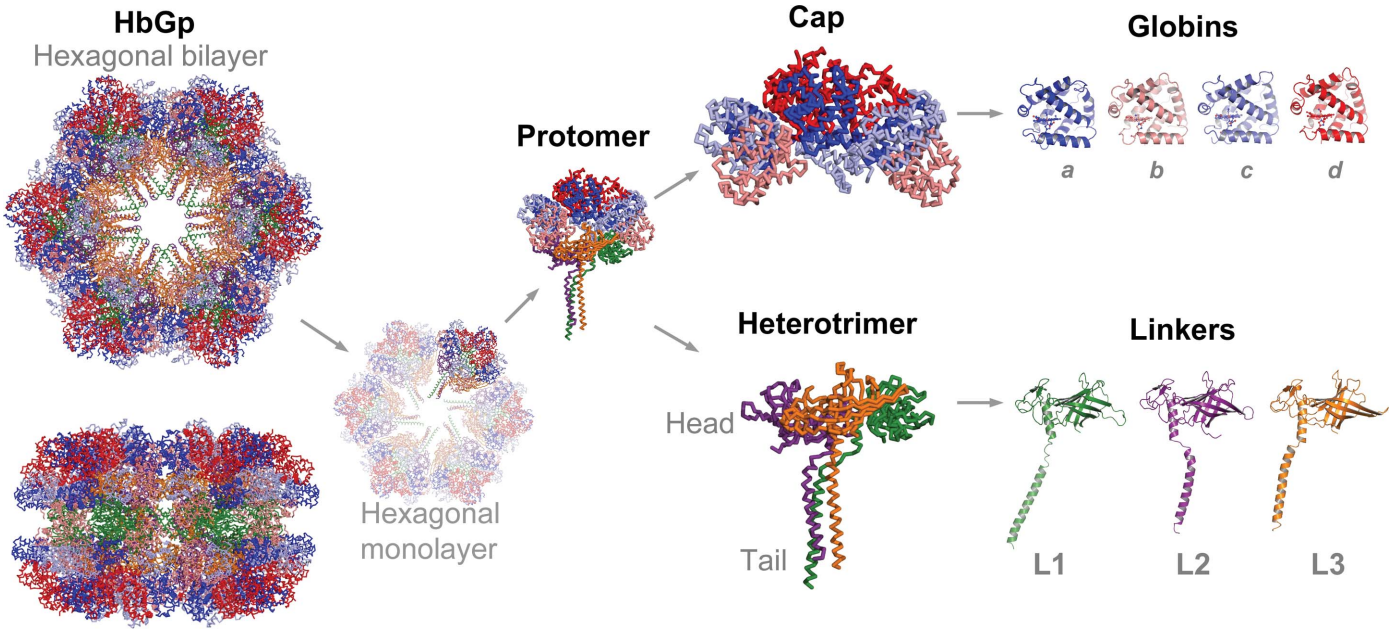


Figure 4
Organization of the 180 polypeptide chains within the particle. To the left of the figure the full particle is shown viewed along both the sixfold axis and one of the twofold axes. A single hexagonal layer (centre left) can be dissected into six protomers composed of a dodecameric globin cap together with a heterotrimeric linker complex. Each globin presents the classical fold based on seven α-helices and each linker is composed of a globular head region together with a tail which points towards the sixfold axis. The three tails form a coiled coil important for particle stability. This figure defines a colour code for the different chains which will be used throughout this paper (top and bottom right).

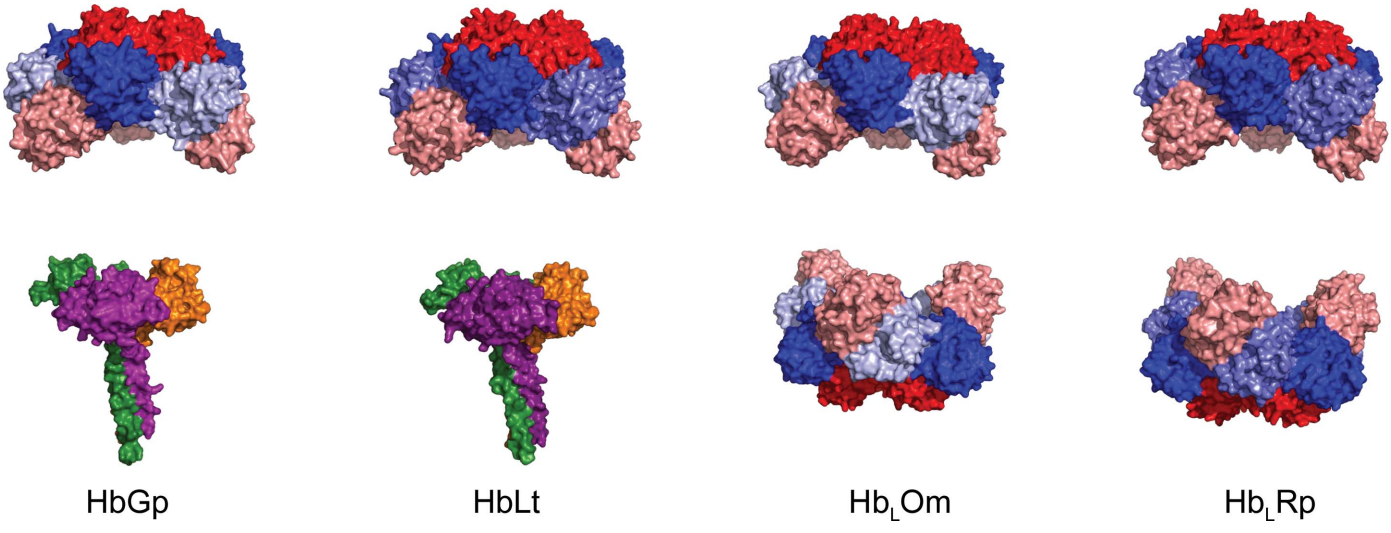


Figure 5
The dodecameric cap. The upper part of the figure shows the similarity between the cap observed in the land-living species (HbGp and HbLt) and half of the hollow spherical structure seen in *O. mashikoi* (Hb$_L$Om) and *R. pachyptila* (Hb$_L$Rp). In the latter a second dodecameric cap completes the 24-mer, whilst in HbGp and HbLt the trimeric complex of linkers interacts in an approximately analogous fashion. In both cases a hollow cavity is generated.

the human counterpart. This is owing to an insertion of three residues within the short loop connecting helices E and F, leading to an N-terminal extension of the latter.

The first level of oligomeric association between globin chains is the generation of *ad* and *bc* dimers. These are equivalent structures, where the interface occurs directly *via* haem groups and residues located in helices E and F (residues Gln98 and His99), as can be seen in Supplementary Fig. S3. The association between a pair of these dimers results in a heterotetramer (Supplementary Fig. S3) in which subunit *a* forms a disulfide bond to subunit *b*. The heterotetramer is repeated three times to form the dodecameric cap. The contacts between heterotetramers that generate the cap occur between subunits a_1 and c_2 (including a disulfide bridge) and between d_1, d_2 and d_3 (where the subscripts refer to different tetramers). The latter contacts result in a trimer of *d* subunits sitting on the threefold axis of symmetry which the cap presents. It is known that the *abc* trimer, when isolated from the entire particle of HbLt, displays a small cooperative behaviour ($C_{\rm Hill} \simeq 1.3$; Fushitani & Riggs, 1991), while the fraction containing the *d* subunit displays no cooperativity. However, when the *abc* trimer is rejoined with the *d* subunit, the resulting cap displays, at pH 6.8, a cooperative behaviour that is indistinguishable from the whole particle ($C_{\rm Hill} = 7.8$). Thus, the structure of the cap is pointed to as being the basic cooperative unit in HbLt. This fact is further supported by the 'light' Hbs (HB_L) found in vestimentiferans and pogonophorans (Negrisolo *et al.*, 2001). Despite the low sequence identity between the subunits in HB_L and those of HbGp and HbLt, the same hierarchal association is preserved, resulting in a dodecameric structure equivalent to the cap, suggestive of a similar cooperative mechanism. In HB_L, two dodecamers join to form a 24-mer with a hollow cavity. The superposition of these dodecamers with the caps of HbGp and HbLt yields r.m.s.d.s of between 1.3 and 1.6 Å on C^{α} atoms, where the orientation of the subunits is effectively identical. Fig. 5 shows a comparison between the structure of the cap in HbGp and the equivalent structures in *L. terrestris*, *O. mashikoi* (Hb_LOm) and *R. pachyptila* (Hb_LRp). The latter two present metal-binding sites (Zn^{2+} in Hb_LRp and Ca^{2+} in Hb_LOm) that are not conserved in the land-living species, indicating that these metals may have a different influence over cooperativity and allostery in these systems.

3.5. Linkers and the heterotrimer

The three types of linker in HbGp show a similar structure divided into four domains (Royer *et al.*, 2006). The first two domains are responsible for forming the coiled coil and consist of two helices connected by a loop. These loops are the most variable regions of the coiled-coil structure when comparing one linker with another. This variation leads to differences in the relative orientations of the two helices in different linkers (Supplementary Fig. S4) and is necessary for producing the inclined orientation of the second coiled-coil region with respect to the first. This variation is conserved on comparing HbGp with HbLt, suggesting that it is relevant to the correct assembly of the trimer (Supplementary Fig. S4). The third domain is cysteine-rich and similar in fold to LDL-A domains, as observed in the LDL receptor. The LDL-A domain contains a Ca^{2+}-binding site formed by six oxygen ligands, forming a distorted octahedral complex. Such a configuration is also compatible with a Mg^{2+} ion, and it is worth noting that Mg^{2+} has a very similar effect on HbGp. However, ICP analysis of the cyano-HbGp sample used in this work for crystallization excluded the presence of magnesium. The fourth and final domain is a typical eight-stranded antiparallel β-barrel which plays an important role in stabilizing the protomer. This globular domain is responsible for the main interactions between the trimer of linkers and the cap.

In the L3 subunits, besides confirmation of the Ca^{2+} site in the LDL-A domain, two other intense peaks were found which were attributed to Zn^{2+} ions based on the coordination environment and ICP analysis (which exclude other metals than Ca and Zn). The first site (zinc site 1) resides between the LDL-A domain and the β-barrel domain. In this case the Zn^{2+} is coordinated by the side chains of His64,

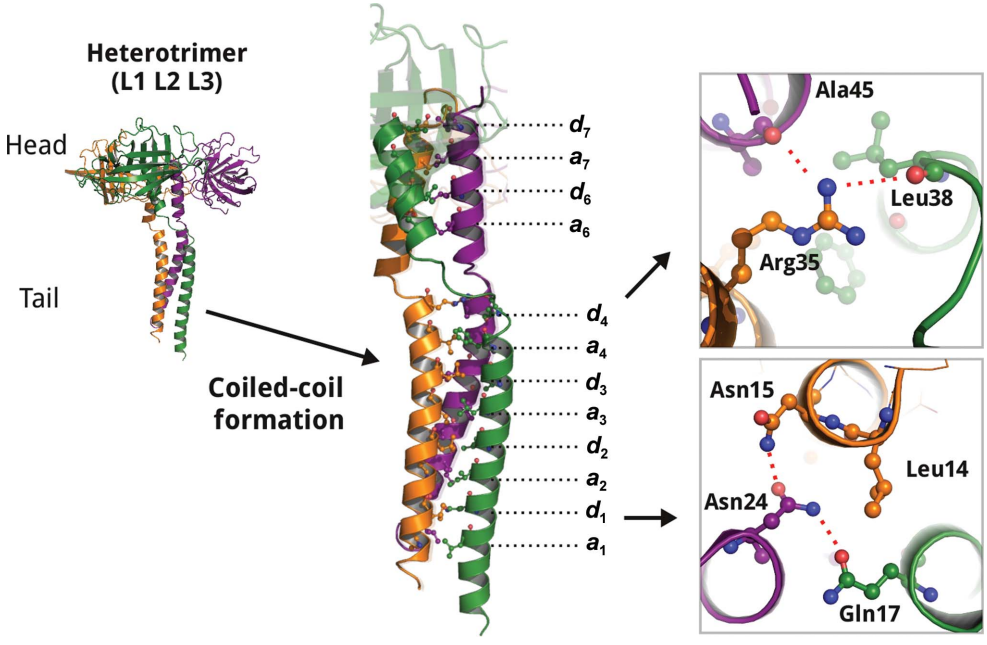


Figure 6
Details of the coiled coil. The coiled coil of the heterotrimer of linkers (left) is shown in detail (centre). The *a* and *d* positions of the heptad repeats are shown explicitly. Details of the d_1 and d_4 layers, which include the presence of hydrophilic residues, are shown explicitly on the right.

Asp90 and His94, and the fourth ligand has been modelled as a water molecule (Fig. 3*b*). Based on the alignment (Fig. 2), this site is present only in HbGp. The second Zn^{2+} site is distant from the first and is found between subunits L3 and L2 from neighbouring protomers of the same hexagonal layer (a fuller description of the interactions made between different protomers will be given later). This site is formed by His142 and His144 from L3, His236 from L2 and a water molecule (Fig. 3*c*). Liochev and coworkers reported superoxide dismutase activity for HbLt, which is in agreement with the presence of Cu^{2+} and Zn^{2+} species as experimentally determined for this system (Liochev *et al.*, 1996). This activity was attributed to the linkers, since the isolated globin portion showed no SOD activity. However, our ICP analysis showed no evidence for copper in HbGp and there is also no evidence in our structure for a bimetallic site as observed in Cu/Zn superoxide dismutases. Our results shed no light on the observations of Liochev and coworkers, which require further investigation and may be a peculiarity of HbLt and not erythrocruorins in general.

The three different linkers interact to form a trimer stabilized principally by hydrophobic interactions from the first two domains, which form the coiled coil. This is directed towards the centre of the structure and is important for particle stability. The coiled coil is interrupted around residue 45 (alignment position 65 in Fig. 2), dividing it into two portions (domains 1 and 2). A full description of the details of the coiled coils and some of their specific features is given in the Supporting Information; some of them are shown in Fig. 6.

3.6. Interactions between the dodecameric caps and linkers

The protomer is formed by interaction between the cap and the trimer of linkers, producing a structure with pseudo-threefold symmetry. The contacts between the cap and linkers occur between the outer rim of the head (globular region) of the linker trimer and the inside edge of the cap, with the LDL-A and β-barrel domains of each linker interacting with one of the globin trimers ($a_1b_1c_2$). As a result of the dome shape of the cap, a large cavity is created within the protomer (Supplementary Fig. S5). The cavity has a volume of around 40 000 Å^3, which is approximately twice the volume occupied by a globin subunit alone. There are five different types of channel which give access to the cavity, giving rise to 13 in total. The first channel has a diameter of approximately 6.5 Å and is unique in that it lies along the threefold axis at the top of the cap formed at the interface between the *d* subunits. The second group corresponds to the channels formed at the interface between neighbouring *abcd* tetramers and has a diameter of approximately 7.0 Å. In the Hb_LOm structure, this channel is occupied by a Ca^{2+}-binding site that is not conserved in HbGp or HbLt. The third group is formed at the interface between the second helical domain, the LDL-A domain and the β-barrel domain of a given linker. The fourth group of channels is formed at the interface between two linkers, specifically between the β-barrel of one linker and the LDL-A domain of its neighbour, resulting in small spaces that give access to the central cavity. The group 3 and group 4 channels have average diameters of 10 Å. The fifth and final group of access channels is the result of interaction between the cap and the heterotrimeric linkers. This channel is formed at the interface between the *bc* globin dimer, the β-barrel from one linker and the LDL-A domain of an adjacent linker. These latter channels are the widest, having diameters of the order of 15 Å. No significant electron density is observed within this cavity, providing no evidence for the presence of small ligands. However, the multiple access channels which arise from defects at subunit or domain interfaces are suggestive of a carrier function for the cavity. However, this is speculation and if the cavity has any physiological significance it remains to be elucidated.

3.7. Interactions between protomers from the same layer

The linker subunits provide the primary contacts between the protomers which generate the full biological particle. The lateral contacts between protomers within the same hexagonal layer occur mainly between L2 of one protomer and L3 from another. The main contacts are hydrogen bonds provided by the antiparallel alignment of β-strand 2 from L2 with β-strand 4 from L3 of the neighbouring protomer. The contact is further stabilized by a salt bridge between Lys113 (L2) and Glu159 (L3). The main difference with respect to HbLt is the presence of zinc site 2, as described above and shown in Fig. 7. The three histidine residues which form the metal-binding site are conserved in other terrestrial species (Fig. 2), suggesting the presence of Zn^{2+} at this interface in all of them.

Zn^{2+} has been described as being different from other divalent cations in terms of its effector properties, at least in some species. According to Ochiai and coworkers, in *Pheretima hilgendorfi* Zn^{2+} binds in a different site from the alkaline-earth metals and acts on the T-state, increasing its affinity for oxygen (Ochiai *et al.*, 1993). It was also shown that zinc is capable of increasing the stability of the particle under conditions which otherwise lead to the fragmentation of the complex, for example at alkaline pH. The conservation of the residues involved at the second zinc site and its strategic location between protomers suggest that it may well be this site which is responsible for contributing to the stability of the particle.

Lateral contacts between protomers also include minor interactions between adjacent caps. Two contacts are observed, the first involving the *a* chain with its counterpart from the neighbouring protomer, and a second involving subunit *b* with subunit *c*. The latter involves the loop connecting helices F and G of the *b* chain with helices G and H from the *c* chain, including a salt bridge between residues Asp102 of subunit *b* and Lys108 of subunit *c*. In summary, there are significant contacts between caps, raising the possibility of communication between globin chains of different protomers.

3.8. Interactions between protomers from different layers

The specific interactions made between protomers from different hexagonal layers occur between those which are effectively superposed (ignoring the small relative rotation of

16° of the two layers with respect to one another which is characteristic of the type I arrangement). These protomers are related by one of the two families of twofold axes present in point group 622. The interaction involves the globular region of linkers L1 together with the helices of the first domain from linkers L2 and L3 (part of the coiled coil). It is worth mentioning that there are no interactions between the caps so that the two hexagonal layers only interact through the linker subunits, and the globin chains in different layers are therefore completely independent. The contact between the β-barrels of the L1 subunits (involving β-strands 1–5) results in a buried surface area of 1102 Å^2 and forms at least 16 hydrogen bonds. The coiled-coil structures intersect at an angle of 100°, and the interface is characterized by a high concentration of charged residues forming ionic interactions: six in total (Fig. 8).

The last interaction between protomers of different hexagonal layers occurs between those which are related by the second family of twofold axes present in point group 622 (D_6). These are made between a given protomer and the neighbour of its superposed partner in the other layer. These contacts are made by their respective coiled coils, which also pack at an angle close to 100° and exclusively involve the L1 subunits. Owing to the symmetry involved, residues Thr12, Asn16 and Gln17 of each subunit end up generating four hydrogen bonds at this interface (Fig. 8).

From the description above, it is clear that the linkers play a key role in stabilizing the full biological particle and particularly in maintaining the two hexagonal discs together. This is emphasized by the fact that no contacts occur between coiled coils of the same layer. The interactions described above lead to each coiled coil being sandwiched between two other coiled coils from adjacent protomers of the opposite layer. This interdigitation creates a network of interactions which appears to lend great stability to the complex and is presumably the principal role of the linkers, depending little or not at all on the globins. This is emphasized by the observation of cap-free hexagonal bilayers composed solely of linkers in electron-microscopy images (data not shown).

3.9. The structure of the *d*-HbGp subunit and cap formation

The structure of the isolated globin *d* chain (*d*-HbGp) was solved independently of the remainder of the structure and presents a global fold very similar to that observed in the full HbGp complex, preserving all of the elements of secondary structure. The isolated *d* chains from both HbLt and HbGp are largely monomeric, although a small fraction is present in the form of oligomers, the presence of which is concentration-dependent (Carvalho, Caravalho *et al.*, 2011; Carvalho, Santiago *et al.*, 2011). Given that three *d* chains come together at the pseudo-threefold axis of the protomer (where each chain forms contacts with its two neighbours), it was imagined that this small oligomeric fraction could be a trimer, which might act as a nucleating agent for cap formation. However, *d*-HbGp crystallizes with a monomer in the asymmetric unit of space group *I*222, precluding the possibility of trimers in the crystal structure. On the other hand, one of the symmetry

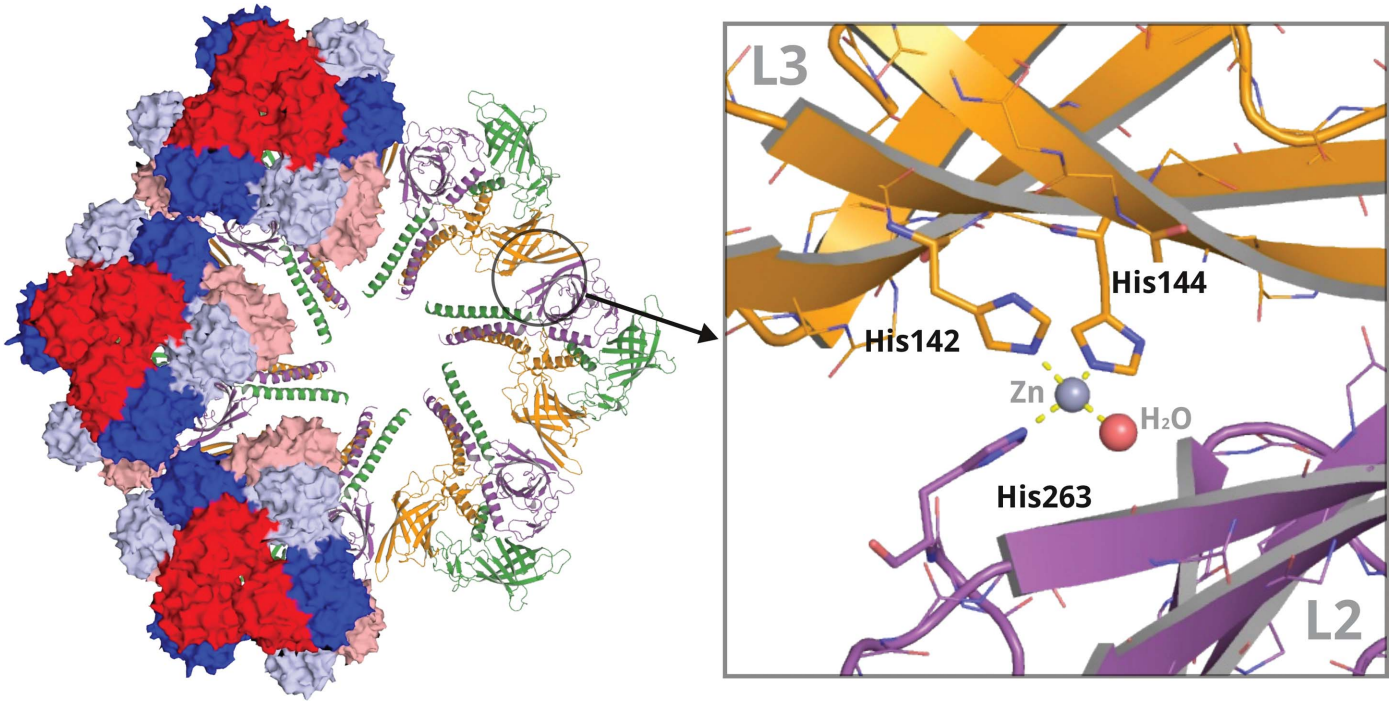


Figure 7
Protomer–protomer interactions within a hexagonal layer. Interactions between the β-barrel domains of linker chains from adjacent protomers contribute to the stability of the hexagonal disc. Hydrogen bonding between antiparallel β-strands from L2 of one protomer and L3 of another together with a previously undescribed metal-binding site (here interpreted as Zn^{2+}) contribute to the interface.

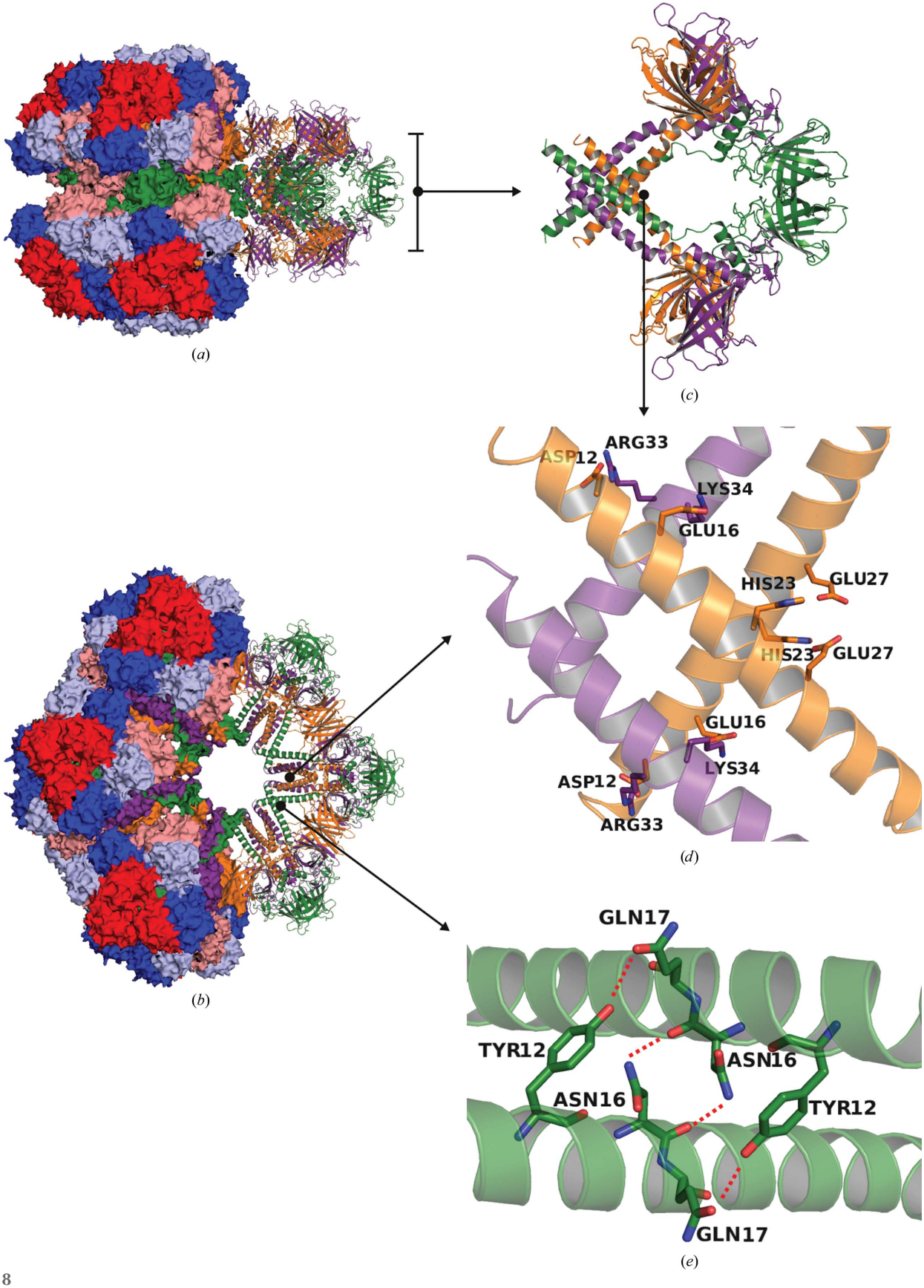


Figure 8
Interactions made between hexagonal layers. (*a*) and (*b*) show overviews of the full particle with part of the molecular surface removed. (*c*) Interactions between layers are made principally by the β-barrel domains of L1 and *via* the coiled coils. (*d*, *e*) Details of the interactions between L2/L3 coiled coils made between 'superimposed' protomers (*d*) and those made by L1 coiled coils between nonsuperimposed protomers (*e*).

mates leads to the generation of a dimer (*dd*) employing an equivalent interface to that made between the *a* and *d* subunits of the full particle (Fig. 9). This interface principally employs the E and F helices, which sandwich the haem group. The measured buried surface area for the *dd* interface is 646.8 Å^2, which is somewhat lower than that presented by the native *ad* interface (914.5 Å^2). Nevertheless, the two interfaces present a similar estimate for the interaction energy (−11.0 and −13.7 kcal mol^{-1}, respectively) as calculated by *PISA* (Krissinel & Henrick, 2007). By contrast, the trimeric structure d_{x3}, present in the entire particle as part of the dodecameric cap, presents an interface between adjacent subunits of only 331.4 Å^2 and an estimated interaction energy of −3.6 kcal mol^{-1}. These data, combined with the interfaces found in the crystal lattice of *d*-HbGp, show that the existence of a trimeric species is not favoured in solution, suggesting that the predominant oligomeric forms are dimers or possibly multiples of dimers.

This is broadly consistent with the experimental evidence from MALDI-TOF MS studies (Oliveira *et al.*, 2007; Carvalho, Caravalho *et al.*, 2011), as well as analytical ultracentrifugation data (Carvalho, Santiago *et al.*, 2011), which show that the isolated *d* chain in solution appears as an equilibrium mixture of monomeric and dimeric species and the contribution of dimers increases at higher protein concentrations. It is also consistent with the original observations of Fushitani & Riggs (1991), which demonstrated that the *d* chains of HbLt also self-assemble into dimers.

The fact that the d_{x3} trimer does not form spontaneously in solution makes it unlikely that this would be an intermediate during cap formation. This would suggest that the fragile native contacts made between the three *d* subunits around the threefold axis in the final assembly are only favoured in the presence of the remaining globin chains. This also suggests that inbuilt negative design disfavouring non-native homodimers (such as *dd*) is not a source of specificity for correct oligomeric assembly, in contrast to that observed in the case of other systems involving a simple competition between homodimers and heterodimers (Bolon *et al.*, 2005). In the present case, the contributions from the formation of the remaining interfaces of the cap would appear to be important in guiding correct assembly.

Since the cap is a stable structure in the absence of the linkers (Strand *et al.*, 2004), this indicates that the assembly of the cap probably involves the formation of *abcd* tetramers, which subsequently associate. This seems more likely than the formation of an $(abc)_3$ intermediate, which would require the subsequent addition of the *d*-chain trimer. It is known that the globin chains of HbLt have the ability to rebuild the cap after fragmentation in alkaline pH. On the other hand, a hexagonal bilayer of HbGp linkers alone has been observed in electron-microscopy experiments (unpublished data), suggesting that this structure could act as a nucleating agent, to which the caps could be added for the formation of the entire complex.

It is of interest to note that the residues which participate in the *ad* interface are largely conserved when comparing the two different isoforms of both the *a* and *d* subunits. However, there is a complementary substitution involving glycines and serines at position 23 of the *a* subunit and position 80 in the *d* subunit, suggesting the possibility of a degree of specificity in which *a*2 may prefer to pair with *d*1 and *a*1 with *d*2. At the *dd* interface of the trimer there is a large accumulation of charged residues, and one salt bridge, between Asp37 of one subunit and Arg34 of its neighbour, appears to be particularly important. Both these residues are substituted by noncharged hydrophilic residues in the *d*2 isoform, suggesting that a mixture of *d*1 and *d*2 within a trimer may be disfavoured.

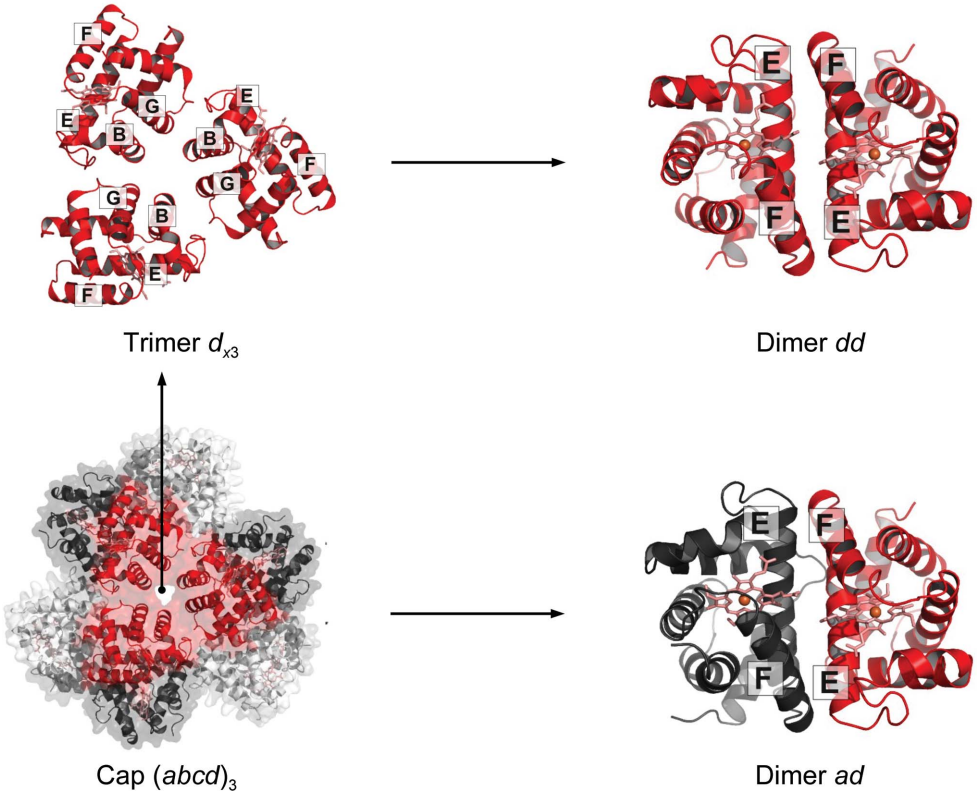


Figure 9
Structure of the isolated *d* chain. The cap, as observed in the full particle, is shown at the bottom left, with the isolated trimer of *d* chains above it. The interface between the *d* and *a* chains observed in the cap is shown at the bottom right; above it (top right) the analogous interactions formed by a *dd* interface observed in the structure of the isolated *d* chain are shown.

4. Conclusions

Giant haemoglobins are remarkable multimeric complexes which display properties related to fundamental questions in biochemistry such as the basis for specificity at protein–protein interfaces, the stability of macromolecular complexes, cooperativity between subunits and spontaneous assembly. There is still a paucity of structural data for these systems and despite its modest resolution, the structure described here is still nominally the highest reported to date. We have described the most important subunit–subunit interfaces within the complex, some for the first time. The reason for the existence of isoforms of certain polypeptide chains remains unclear, as is the extent to which the assembled particles may be heterogeneous. The difficulty in understanding specificity at interfaces is highlighted by the structure of the isolated *d* chain, which uses a promiscuous interface to form a homodimer in place of the trimer observed at the local threefold axis of the full particle. However, this unexpected result sheds some light on how the complex might be assembled *in vivo*, as it eliminates the possibility of the trimer serving as a nucleus for the formation of the cap. With respect to the stability of the complex, the discovery of an interfacial zinc ion between different protomers seems to be particularly relevant as it is in agreement with previous reports describing the impact of zinc at low concentrations on the disassembly of the complex. The conservation of the residues which form the binding site suggests this zinc ion to be an important structural component of all such erythrocruorins. On the other hand, a previously reported zinc site in HbLt should be treated with caution as its environment seems to be more compatible with an anion, the physiological relevance of which remains to be elucidated. It is to be expected that the intersubunit contacts described in detail for HbGp in this paper will also allow a better understanding of the main driving forces leading to several experimental observations that at present lack an explanation at the molecular level, such as the oligomeric dissociation of the complex under alkaline conditions. Most importantly, there is still very little known about the structural mechanism behind the cooperativity and the elevated Hill coefficient that erythrocruorins present. The structure reported here and that of HbLt are good starting points, but future efforts must focus on attempts to produce crystals of well defined R-states and T-states of HbGp.

5. Related literature

The following references are cited in the Supporting Information for this article: Harbury *et al.* (1993), Ownby *et al.* (1993) and Richardson & Richardson (1988).

Acknowledgements

We gratefully acknowledge a CNPq fellowship awarded to JFRB. We are also grateful to Sandra Martha Gomes Dias of the National Laboratory of Biosciences (LNBio) for useful discussions concerning nucleic acid sequencing. We acknowledge Dr Maria Inês Basso Bernardi (IFSC/USP) for the determination of heavy atoms in the sample using inductively coupled plasma atomic emission analysis and Dr Annie Héroux for her contributions during diffraction data collection at NSLS beamline X25C. This work was financed by FAPESP and CNPq

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Bachega, J. F. R., Bleicher, L., Horjales, E. R., Santiago, P. S., Garratt, R. C. & Tabak, M. (2011). *J. Synchrotron Rad.* **18**, 24–28.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 12724–12729.
Bosch Cabral, C., Imasato, H., Rosa, J. C., Laure, H. J., da Silva, C. H. T. de P., Tabak, M., Garratt, R. C. & Greene, L. J. (2002). *Biophys. Chem.* **97**, 139–157.
Carvalho, F. A. O., Carvalho, J. W. P., Alves, F. R. & Tabak, M. (2013). *Int. J. Biol. Macromol.* **59**, 333–341.
Carvalho, F. A. O., Carvalho, J. W. P., Santiago, P. S. & Tabak, M. (2011). *Proc. Biochem.* **46**, 2144–2151.
Carvalho, F. A. O., Santiago, P. S., Borges, J. C. & Tabak, M. (2009). *Anal. Biochem.* **385**, 257–263.
Carvalho, F. A. O., Santiago, P. S., Borges, J. C. & Tabak, M. (2011). *Int. J. Biol. Macromol.* **48**, 183–193.
De Haas, F., Biosset, N., Taveau, J.-C., Lambert, O., Vinogradov, S. N. & Lamy, J. N. (1996). *Biophys. J.* **70**, 1973–1984.
De Haas, F., Taveau, J.-C., Boisset, N., Lambert, O., Vinogradov, S. N. & Lamy, J. N. (1996). *J. Mol. Biol.* **255**, 140–153.
De Haas, F., Zal, F., You, V., Lallier, F., Toulmond, A. & Lamy, J. N. (1996). *J. Mol. Biol.* **264**, 111–120.
Elmer, J. & Palmer, A. F. (2012). *J. Funct. Biomater.* **3**, 49–60.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Flores, J. F., Fisher, C. R., Carney, S. L., Green, B. N., Freytag, J. K., Schaeffer, S. W. & Royer, W. E. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 2713–2718.
Fushitani, K. & Riggs, A. F. (1991). *J. Biol. Chem.* **266**, 10275–10281.
Grabherr, M. G. *et al.* (2011). *Nature Biotechnol.* **29**, 644–652.
Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). *Science*, **262**, 1401–1407.
Harnois, T., Rousselot, M., Rogniaux, H. & Zal, F. (2009). *Artif. Cells Blood Substit. Immobil. Biotechnol.* **37**, 106–116.
Hirsch, R. E., Jelicks, L. A., Wittenberg, B. A., Kaul, D. K., Shear, H. L. & Harrington, J. P. (1997). *Artif. Cells Blood Substit. Immobil. Biotechnol.* **25**, 429–444.
Kapp, O. H., Vinogradov, S. N., Ohtsuki, M. & Crewe, A. V. (1982). *Biochim. Biophys. Acta*, **704**, 546–548.
Kleywegt, G. & Jones, T. A. (2002). *Structure*, **10**, 465–472.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Liochev, S. I., Kuchumov, A. R., Vinogradov, S. N. & Fridovich, I. (1996). *Arch. Biochem. Biophys.* **330**, 281–284.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Negrisolo, E., Pallavicini, A., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A., Moens, L. & Lanfranchi, G. (2001). *J. Biol. Chem.* **276**, 26391–26397.
Numoto, N., Nakagawa, T., Kita, A., Sasayama, Y., Fukumori, Y. & Miki, K. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 14521–14526.
Numoto, N., Nakagawa, T., Kita, A., Sasayama, Y., Fukumori, Y. & Miki, K. (2008). *Biochemistry*, **47**, 11231–11238.

Ochiai, T., Hoshina, S. & Usuki, I. (1993). *Biochim. Biophys. Acta*, **1203**, 310–314.
Oliveira, M. S., Moreira, L. M. & Tabak, M. (2007). *Int. J. Biol. Macromol.* **40**, 429–436.
Ownby, D. W., Zhu, H., Schneider, K., Beavis, R. C., Chait, B. T. & Riggs, A. F. (1993). *J. Biol. Chem.* **268**, 13539–13547.
Perutz, M. (1997). *Unravelling The Atomic Mechanism of Haemoglobin*. Singapore: World Scientific.
Richardson, J. S. & Richardson, D. C. (1988). *Science*, **240**, 1648–1652.
Rousselot, M., Delpy, E., Drieu La Rochelle, C., Lagente, V., Pirow, R., Rees, J.-F., Hagege, A., Le Guen, D., Hourdez, S. & Zal, F. (2006). *Biotechnol. J.* **1**, 333–345.
Royer, W. E., Omartian, M. N. & Knapp, J. E. (2007). *J. Mol. Biol.* **365**, 226–236.
Royer, W. E., Sharma, H., Strand, K., Knapp, J. E. & Bhyravbhatla, B. (2006). *Structure*, **14**, 1167–1177.
Royer, W. E., Strand, K., van Heel, M. & Hendrickson, W. A. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 7107–7111.
Santiago, P. S., Carvalho, F. A. O., Domingues, M. M., Carvalho, J. W. P., Santos, N. C. & Tabak, M. (2010). *Langmuir*, **26**, 9794–9801.
Santiago, P. S., Carvalho, J. W. P., Domingues, M. M., Santos, N. C. & Tabak, M. (2010). *Biophys. Chem.* **152**, 128–138.
Standley, P. R., Mainwaring, M. G., Gotoh, T. & Vinogradov, S. N. (1988). *Biochem. J.* **249**, 915–916.
Strand, K., Knapp, J. E., Bhyravbhatla, B. & Royer, W. E. (2004). *J. Mol. Biol.* **344**, 119–134.
Vinogradov, S. N. (1985). *Comp. Biochem. Physiol. B*, **82**, 1–15.
Weber, R. E. & Vinogradov, S. N. (2001). *Physiol. Rev.* **81**, 569–628.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xie, Q., Donahue, R. A., Schneider, K., Mirza, U. A., Haller, I., Chait, B. T. & Riggs, A. F. (1997). *Biochim. Biophys. Acta*, **1337**, 241–247.