Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The structure of a class 3 nonsymbiotic plant haemoglobin from *Arabidopsis thaliana* reveals a novel N-terminal helical extension

Brandon J. Reeder* and Michael A. Hough

School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England

Correspondence e-mail: reedb@essex.ac.uk

Received 26 November 2013
Accepted 3 March 2014

PDB reference: AHb3, 4c0n

Plant nonsymbiotic haemoglobins fall into three classes, each with distinct properties but all with largely unresolved physiological functions. Here, the first crystal structure of a class 3 nonsymbiotic plant haemoglobin, that from *Arabidopsis thaliana*, is reported to 1.77 Å resolution. The protein forms a homodimer, with each monomer containing a two-over-two α-helical domain similar to that observed in bacterial truncated haemoglobins. A novel N-terminal extension comprising two α-helices plays a major role in the dimer interface, which occupies the periphery of the dimer–dimer face, surrounding an open central cavity. The haem pocket contains a proximal histidine ligand and an open sixth iron-coordination site with potential for a ligand, in this structure hydroxide, to form hydrogen bonds to a tyrosine or a tryptophan residue. The haem pocket appears to be unusually open to the external environment, with another cavity spanning the entrance of the two haem pockets. The final 23 residues of the C-terminal domain are disordered in the structure; however, these domains in the functional dimer are adjacent and include the only two cysteine residues in the protein sequence. It is likely that these residues form disulfide bonds *in vitro* and it is conceivable that this C-terminal region may act in a putative complex with a partner molecule *in vivo*.

1. Introduction

Plants contain several nonsymbiotic haemoglobins (nsHbs) that are distinct in properties and function from their well characterized mammalian homologues. Understanding the structure and physiological function of these nsHbs has been a goal for many years. *Arabidopsis thaliana* expresses three nsHbs, classed as AHb1, AHb2 and AHb3. The first two classes of globins are defined from their relative affinities to bind oxygen (Smagghe *et al.*, 2009), whereas class 3 nsHbs resemble the truncated globins found in bacteria (Smagghe *et al.*, 2009; Watts *et al.*, 2001). The physiological functions of these proteins are still under debate; however, AHb1 exhibits a high NO dioxygenase activity and may play a role in NO regulation (Dordas *et al.*, 2003; Hill, 2012; Perazzolli *et al.*, 2004), with distal hydrophobic cavities supporting a role in NO binding (Bruno *et al.*, 2007). Other possible functions include enhanced tolerance to peroxide stress (Yang *et al.*, 2005). Various hypotheses have been presented for the function of AHb2, ranging from oxygen binding to NO dioxygenase and peroxidase activities, although ligand-binding properties support a distinctive physiological role compared with AHb1 (Bruno *et al.*, 2007).

Class 3 nsHbs (also referred to as GLB3) appear to be ubiquitous in plants (Garrocho-Villegas *et al.*, 2007; Hunt *et al.*, 2001; Smagghe *et al.*, 2009), yet there has been little evidence to date for their potential physiological functions. AHb3 is

widely distributed in *Arabidopsis* roots and shoots and is downregulated by hypoxia (Dordas *et al.*, 2003; Watts *et al.*, 2001). Previous studies have characterized CO and O_2 binding to AHb3 through stopped-flow and flash-photolysis kinetics (Watts *et al.*, 2001). Although AHb3 showed a linear relationship between ligand concentration and binding kinetics, rebinding following photo-dissociation was unusual, showing kinetics that were independent of CO concentration. AHb3 exhibits a reported transient hexacoordinate iron-ligation state upon reduction of the ferric protein to ferrous protein (Watts *et al.*, 2001).

The α-helical sandwich structures of the haemoglobin superfamily typically have a three-over-three fold which forms the hydrophobic pocket in which the haem moiety resides. This basic structural composition is observed in animal globins such as erythrocyte haemoglobin (Perutz, 1960), myoglobin (Kendrew *et al.*, 1960), neuroglobin (Pesce *et al.*, 2004) and cytoglobin (de Sanctis *et al.*, 2004). In plants these structures are also observed in symbiotic leghaemoglobins and class 1 nsHbs, with a predicted three-over-three structure in class 2 nsHbs. Truncated haemoglobins are a distinct class of globins found mainly in bacteria, are 20–40 residues shorter than other globins (Pesce *et al.*, 2000) and exhibit a two-over-two α-helical conformation. This is usually expressed as a deletion of α-helices, creating a two-over-two α-helical sandwich.

Class 3 nsHbs have a close sequence homology to bacterial truncated Hbs (Pesce *et al.*, 2000, 2007), but the overall sequence length of AHb3 is 175, some 42 residues longer than the truncated haemoglobin from *Bacillus subtilis* and 15–17 residues longer than the three-over-three structures of other class 1 and 2 haemoglobins. The structure and function of the extended N-terminal and C-terminal domains of AHb3 are unknown, but the N-terminal section has a predicted α-helical structure (Watts *et al.*, 2001).

Here, we report the first structure of a class 3 nsHb, that from *Arabidopsis thaliana*. The protein tertiary structure shows a two-over-two α-helical fold and haem-pocket architecture typical of bacterial truncated globins. The protein is a homodimer with a novel N-terminal extension and a dimeric interface structure. This interface consists of a network of hydrogen bonds with a central large open cavity, and a histidine residue appears to function as an inter-subunit anchor. The C-terminal domains are disordered but are adjacent in the dimeric structure and contain cysteine residues of which one appears to be highly conserved. The haem-pocket architecture, while similar to that of truncated globins, has an unusually open pocket structure which may account for the previously reported unusual ligand-binding kinetics.

2. Material and methods

2.1. Cloning, expression and purification of recombinant AHb3

The cDNA for AHb3 (synthesized by Epoch Life Science Inc., Missouri, USA) was subcloned into pET-28a (Novagen) using *Nde*I and *Eco*RI restriction sites such that a cleavable His-tag sequence was added to the N-terminal protein sequence. Plasmids were transformed into *Escherichia coli* BL21 (DE3) cells (Invitrogen) and grown in Luria–Bertani medium at 37°C and 180 rev min^{-1} with 50 µg ml^{-1} kanamycin sulfate. When the cell optical density at 600 nm reached ~1, 500 µ*M* isopropyl β-D-1-thiogalactopyranoside (Fisher) was added to initiate protein expression. Additionally, 250 µ*M* 5-aminolevulinic acid (Organix, Essex, England) and 100 µ*M* ferric citrate (Sigma–Aldrich, Poole, England) were added to facilitate haem synthesis and CO was bubbled through the broth for approximately 1 min before the vessel was sealed. The cells were incubated for a further 18 h at 310 K and 120 rev min^{-1}, whereupon the cells were harvested by centrifugation (11 000*g*, 5 min, 4°C). Following a freeze–thaw cycle, the cells were disrupted using an Avestin C5 EmulsiFlex at ~100 MPa. The disrupted cells were centrifuged (22 000*g*, 30 min, 4°C) to remove cellular debris and the protein was purified using a His-tag nickel-affinity column (GE Healthcare) as per the manufacturer's instructions. Imidazole from the purification procedure was removed by dialysis with 1 m*M* sodium tetraborate pH 9.5 (three changes) followed by incubation with thrombin (Sigma–Aldrich, 10 U per milligram of protein) overnight at 25°C for proteolytic removal of the tag from the protein. The nickel-affinity column was again used to separate the tag and the protein was dialysed and concentrated using a 5000 Da molecular-weight cutoff spin filter.

2.2. Optical properties, extinction coefficients and acid–alkaline transition

The optical spectrum of the ferric protein (~10 µ*M* in 25 m*M* sodium phosphate buffer pH 7.4) was measured using a Varian Cary 5E spectrophotometer. An aliquot of the protein solution was taken and analysed by an Agilent 1100 HPLC using a Zorbax SB300C3 column as described previously (Reeder *et al.*, 2007). The HPLC separates the haem (iron protoporphyrin IX) from the protein, the concentration of which was determined by comparison with horse myoglobin, where the concentration of the deoxy ferrous haem protein was determined optically at 425 nm using an extinction coefficient of 121 mM^{-1} cm^{-1} (Antonini & Brunori, 1971). The remaining AHb3 solution was reduced to the deoxy ferrous oxidation state using sodium dithionite (~10 m*M*) and the optical spectrum was recorded. Approximately 5 ml of CO gas was bubbled through the solution to obtain the ferrous–CO form of the protein. Extinction coefficients of the Soret peak maxima were calculated using the concentration of haem determined by the HPLC measurements.

The optical spectrum of 10 µ*M* AHb3 (5 m*M* sodium acetate, 5 m*M* sodium phosphate and 5 m*M* sodium borate) was taken following measurement of the pH using a calibrated Hanna pH meter connected to a micro pH electrode. The initial pH was 6.0. Sequential additions of 5 µl sodium hydroxide (between 0.25 and 2 *M*) were added to the cuvette, mixed and the pH was determined before the spectrum was recorded. This was repeated until the pH was greater than 10.

Table 1
Crystallographic data-collection and processing statistics.

Values in parentheses refer to the outermost resolution shell. For the unit-cell parameters, note that $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$.

Data set	High resolution	SAD peak
Wavelength (Å)	0.9686	1.7372
Resolution (Å)	1.77	2.19
Space group	$P4_332$	$P4_332$
Unit-cell parameter (Å)	123.4	123.2
Unique reflections	31454	17071
Completeness (%)	99.0 (99.9)	100 (100)
R_{merge} (%)	0.053 (0.768)	0.085 (0.565)
Mean $I/\sigma(I)$	16.4 (1.9)	26.0 (9.4)
Redundancy	5.4 (5.5)	60.0 (56)
Anomalous redundancy	—	32.6 (29.5)
R_{cryst}	0.186	—
R_{free}	0.205	—
ESU based on maximum likelihood (Å)	0.055	—
R.m.s.d., bond lengths (Å)	0.018	—
R.m.s.d., bond angles (°)	1.74	—
Ramachandran favoured (%)	99.3	—
Wilson B factor (Å^2)	24.3	31.2
PDB code	4c0n	—

Optical spectra were corrected for dilution and the changes in the spectra (424–406 nm) were fitted to the Henderson–Hasselbalch equation to obtain the pK using the least-squares method in the *Microsoft Excel Solver* program.

2.3. Crystallization and structure determination

Crystals were grown by the hanging-drop vapour-diffusion method at 20°C. 1 µl 25 mg ml^{-1} protein solution was mixed with an equal volume of reservoir solution consisting of 0.1 *M* Tris, 1.6 *M* ammonium sulfate pH 8.2. Crystals of dimensions ~0.5 × 0.2 × 0.2 mm grew within one week and were transferred to cryoprotectant solution consisting of mother liquor and 25% glycerol before flash-cooling to 100 K by plunging them into liquid nitrogen. High-resolution data were measured to 1.77 Å resolution on Diamond Light Source beamline I24 using a Pilatus 6M detector (Dectris) and an X-ray wavelength of 0.9686 Å. All data were indexed using *iMosflm* (Battye *et al.*, 2011) and were scaled and merged using *SCALA* (Evans, 2006) in the *CCP*4 suite. The structure was solved by SAD phasing using the anomalous signal from the intrinsic haem Fe. High-redundancy anomalous data were measured at the peak of the Fe *K* edge (λ = 1.7372 Å). One Fe site was located per asymmetric unit and the resulting electron-density map allowed an initial model to be automatically built using *Buccaneer* (Cowtan, 2006). This starting model was refined against the 1.77 Å resolution data set using

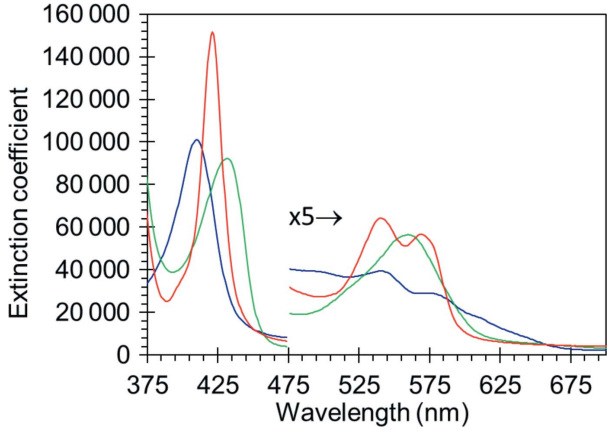

Figure 1
Optical characteristics and extinction coefficients of AHb3 in the ferric (blue), deoxy ferrous (green) and ferrous CO-bound (red) states.

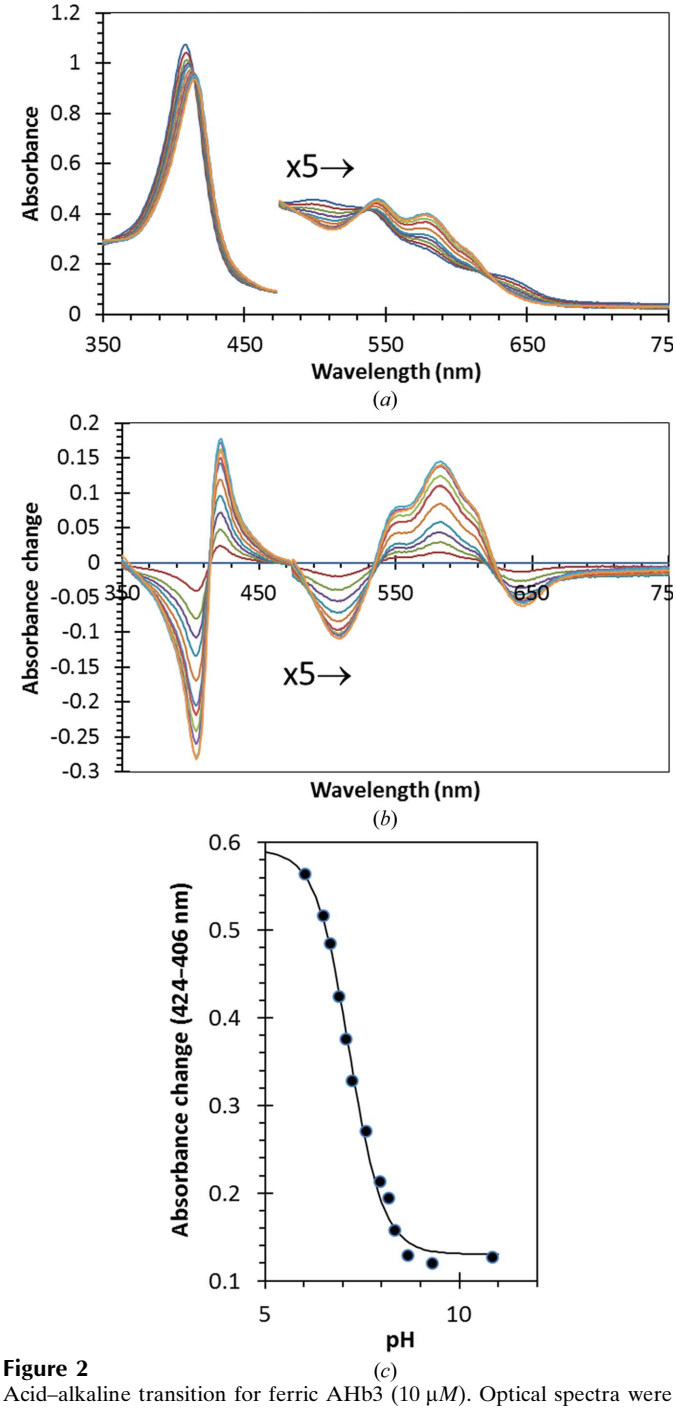

Figure 2
Acid–alkaline transition for ferric AHb3 (10 µ*M*). Optical spectra were measured as a function of pH as absolute (*a*) or difference spectra (*b*) where the low-pH ferric spectrum was made to be silent. (*c*) The measured pK between the aqua and hydroxyl forms (424–406 nm) was 7.17 ± 0.10.

*REFMAC*5 (Murshudov *et al.*, 2011). Riding H atoms were added when refinement of the protein atoms had converged. Models were rebuilt between cycles of refinement in *Coot* and were validated using the *MolProbity* server (Chen *et al.*, 2010) and tools in *Coot* (Emsley & Cowtan, 2004). Coordinates and structure factors were deposited in the RCSB Protein Data Bank (PDB) as entry 4c0n. A summary of the data and refinement statistics and the quality indicators for the structure are given in Table 1.

2.4. Sequence alignments

Protein sequences were obtained from the UniProt website (http://www.uniprot.org) or the PDB (http://www.rcsb.org) and were aligned using *ClustalX*2 (Larkin *et al.*, 2007).

3. Results

3.1. Optical properties of AHb3

Circular dichroism of the isolated recombinant AHb3 showed an α-helical secondary structure typical of haemoglobin proteins, indicating that the protein was properly folded (data not shown). The optical spectrum of AHb3 in the ferric oxidation state showed a Soret peak at 410 nm (Fig. 1), essentially identical to that previously reported (Watts *et al.*, 2001). The extinction coefficient for the ferric protein was determined to be 101 mM^{-1} cm^{-1} at 410 nm and pH 7.4. Additional peaks at 502, 551 and 585 nm and a small shoulder at ∼630 nm are suggestive of protein in a mixture of aqua and hydroxide ferric ligand states similar to that observed in myoglobin and animal erythrocyte haemoglobins (Antonini & Brunori, 1971). This mixture of ligand species was confirmed through a pH titration of the ferric protein (Fig. 2). The optical spectrum exhibited an acid–alkaline transition, shifting the spectrum between the aqua ferric species (Fe^{3+}–H_2O) at low pH with optical bands at 409, 502, 551, 585 and ∼630 nm to a hydroxide form (Fe^{3+}–OH^-) at high pH with the Soret peak at 416 nm, with loss of the 502 and ∼630 nm bands. The 551 and 585 nm bands are much more prominent, with an additional small band at ∼615 nm. The pK between the species (Fig. 2*c*) was measured as 7.17 ± 0.10. It should be noted that below pH 6.5 there was significant optical scattering owing to protein aggregation, which was largely reversible by re-alkalinization.

The dithionite-reduced ferrous spectrum shows a Soret peak at 428 nm ($\varepsilon_{428\ nm}$ = 92 mM^{-1} cm^{-1}) and a single visible peak at 560 nm, characteristic of haem iron in a penta-coordinated state. The ferrous–CO form of the protein shows a Soret peak at 421 nm ($\varepsilon_{421\ nm}$ = 152 mM^{-1} cm^{-1}) with β and α bands at 541 and 570 nm, respectively. It was previously reported that pentacoordination of the deoxy ferrous haem iron was only observed following a hexacoordinate intermediate (Watts *et al.*, 2001). This transition between hexacoordinate and pentacoordinate states in the ferrous protein followed a time course of 10–15 min. To replicate this, stopped-flow spectroscopy was used to follow the optical changes of the protein when reduced by dithionite, following the initial rapid reduction and subsequent optical changes

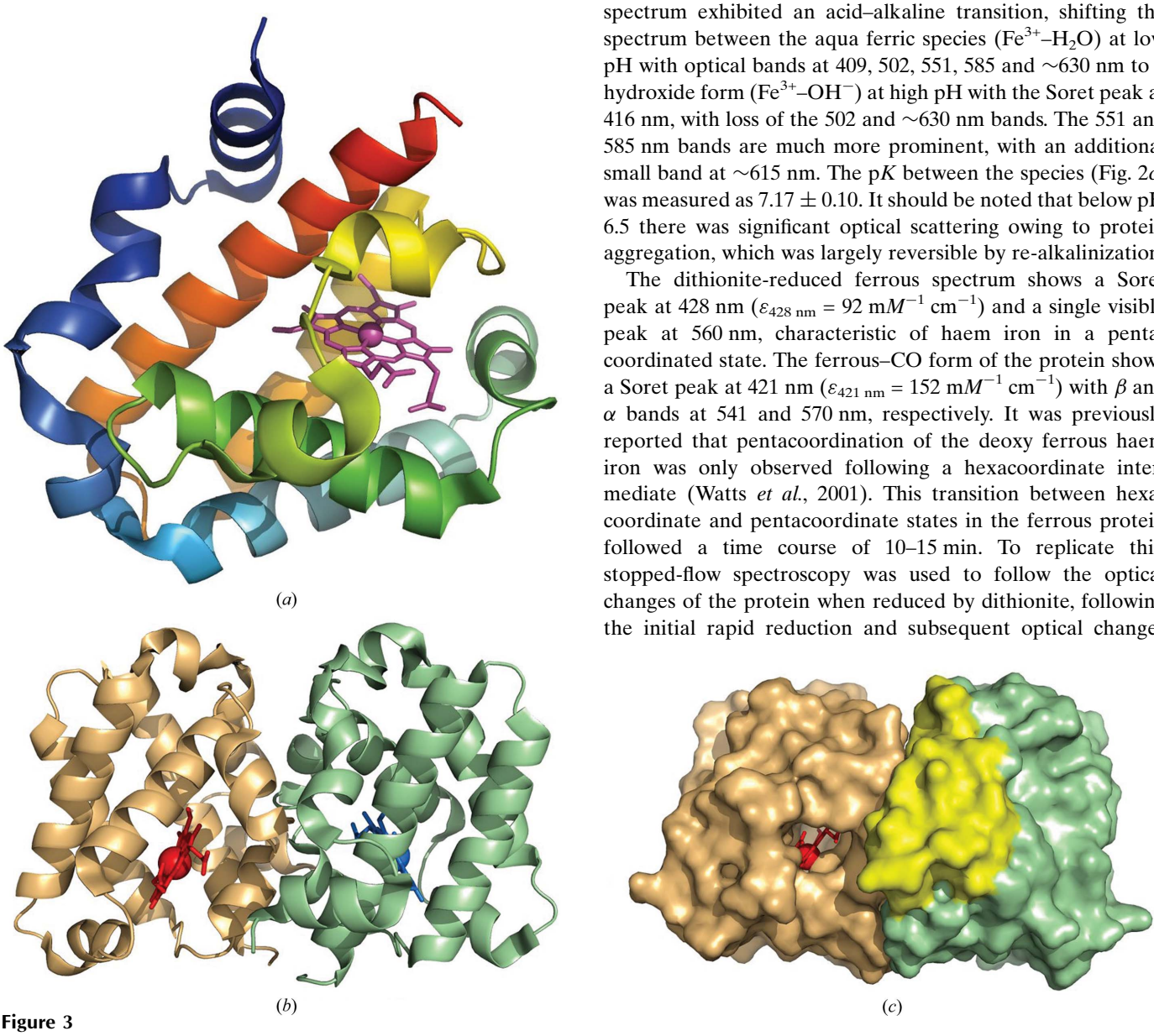

Figure 3
(*a*) Overall protein fold of an AHb3 subunit. The two helices representing the N-terminal region are shown in dark blue and the haem group is indicated in pink. (*b*) The functional dimer. (*c*) Surface representation of the dimer with residues 1–24 at the N-terminus coloured yellow.

Table 2
Key bond lengths and interatomic distances in the dimer interface and haem pocket.

	Bond	Distance (Å)
Dimer	Gln5 $O^{\varepsilon1}$—Arg97 $N^{\eta1}$	2.84
	Gln16 $N^{\varepsilon2}$—Gln87 O†	2.65
	Gln16 $N^{\varepsilon2}$—Ser86 O†	2.65
	His91 $N^{\varepsilon2}$—Thr140 $O^{\gamma1}$	2.81
Haem	Fe—His $N^{\varepsilon2}$	2.10
	Fe—O (OH)	1.85
	OH—Trp111 $N^{\varepsilon1}$	2.88
	OH—Tyr44 O^{η}	3.47
	OH—Gln71 $N^{\varepsilon2}$	4.63

† Only one of two alternate conformations of Gln16 participates in this interaction.

over a 15 min time period (Supplementary Fig. S1). The ferric protein was reduced within 1 s using 5 m*M* dithionite (rate constant 3.4 s^{-1}). The optical spectrum of the initial ferrous spectrum shows that the protein is reduced directly to a species with a 560 nm peak in the visible region, characteristic of a pentacoordinate state, with no evidence of a hexacoordinate intermediate. Subsequent minor optical changes are observed indicating either minor conformational changes in the haem pocket or minor side reactions such as the production of peroxide, which is often observed using dithionite in an oxygenated solution over this time scale (Dalziel & O'Brien, 1957).

3.2. Overall structure of AHb3

The structure of AHb3 was solved to a resolution of 1.77 Å (Fig. 3). Residues 2–151 of the protein sequence were modelled into the electron density, with the remaining 24 residues at the C-terminus presumed to be disordered. The average *B* factor was 25.3 $Å^2$ for main-chain atoms and 27.4 $Å^2$ for side-chain atoms. There is one monomer in the crystallographic asymmetric unit, but from examination of symmetry-related molecules and analysis using *PISA* the biological assembly was identified to be a dimer with a buried surface area of 4570 $Å^2$. The protein displays a two-over-two core fold, similar to that found in bacterial haemoglobins and predicted previously for AHb3 (Pesce *et al.*, 2000; Watts *et al.*, 2001). However, a novel N-terminal extension is observed containing two short helices that lie almost perpendicular to each other and wrap around the surface of the protein (Fig. 3*a*). The first helix comprises residues 3–13, with the second helix comprising residues 15–25. The second helix appears to be strained, with a significant distortion in orientation between the first and second turn of the helix.

3.3. The dimeric interface

The dimeric structure is shown in Figs. 3(*b*) and 3(*c*), with the two haem irons some 23.5 Å apart. The most noticeable feature is that the novel N-terminal domain comprising the first two helices almost exclusively forms the interface between the two homodimers (Fig. 3*c*, shown in yellow). Unusually, the haem group can be observed within the structure of the molecular-surface model of the protein. This indicates a deep solvent cavity extending from the surface to the haem pocket, suggesting that exogenous ligands will be able to readily access the haem pocket from bulk solvent.

The interface between the two monomers consists of a hydrogen-bonding network (Fig. 4*a*). The extended N-terminal region contains two glutamine residues that hydrogen bond to the second protein subunit. A hydrogen bond between the acid group of Gln5 and Arg97 shows a bond length of 2.84 Å (Table 2). Additionally, the amine group of

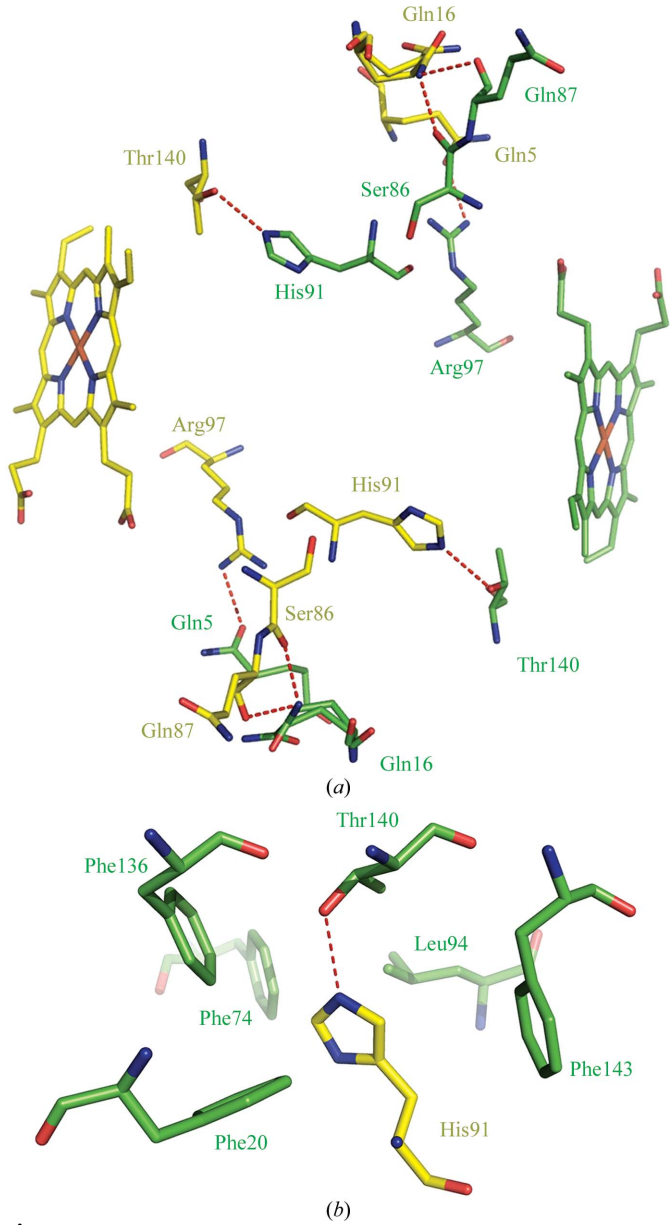

Figure 4
Intersubunit interface structure. (*a*) Hydrogen-bonding network between AHb3 subunits at the dimeric interface. Note the anchoring role of residue His91. (*b*) The intersubunit 'histidine anchor'. His91 from one monomer (yellow) lies within a largely hydrophobic cleft in the second monomer (green residues) and forms a 2.9 Å hydrogen bond to Thr140 (dashed red line). There are two such interaction sites in the functional dimer.

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: QH5004).

Gln16 forms hydrogen bonds to the carbonyl O atoms of Ser86 and Gln87 (both 2.65 Å). Residue His91 extends into a cleft, burying this residue deep within the opposing subunit, apparently interlocking the two subunits. His91 participates in an intersubunit hydrogen bond to Thr140 (2.81 Å; Fig. 4*b*), while the remainder of the cleft is made up of hydrophobic amino acids. Intriguingly, His91 is only 5.8 Å from the nearest atom of the haem from the other monomer of the functional dimer.

The interface between the subunits mainly occurs at the external solvent-facing surface. There is no interaction in the central area of the protein interface, creating a central area with a large cavity that is open to the bulk solvent (Supplementary Fig. S2, shown in green). Note that this cavity is not in the vicinity of the disordered residues at the C-terminus. This cavity is up to 8 Å wide by 18 Å in length, with a volume of ∼700 Å^3. Additionally, there is a smaller cavity below the main cavity (Supplementary Fig. S2, shown in yellow) that appears to link the two open haem-pocket access routes. A strong electron-density peak corresponding to a single atom was present at the dimer interface. This lies in a position to make six weak (∼3.2–3.7 Å) interactions with backbone amines from each monomer. Such coordination is not consistent with any metal ion but is plausible for a chloride, which we have thus modelled (Supplementary Fig. S3). Analysis of the surface of each monomer suggests that the chloride binding pocket is not accessible to bulk solvent and so we propose that binding occurs during protein folding and dimerization.

3.4. Haem pocket

The haem group sits within a pocket that is connected to the protein surface *via* a cavity. The haem iron is coordinated by one proximal protein ligand, His98, with a Fe—N bond length of 2.1 Å (Fig. 5, Table 2). The distal binding site is occupied by a hydroxide molecule at a distance of 1.9 Å, although explicit identification of the ligand is difficult at the resolution of the crystal structure. However, under the conditions of crystallization (pH 8.0), the sixth coordination site of the haem iron is predicted to be predominantly (87%) in the low-spin state with a hydroxide ligand, based on the optical acid–alkaline transition (Fig. 2). Several well ordered water molecules are present within the distal pocket (Fig. 5).

The distal pocket does not contain any His residue that could bind to either Fe or to a Fe-bound water or gas ligand. The pocket is lined by large polar or hydrophobic residues, specifically Phe43, Tyr44, Phe59, Gln71, Phe75 and Trp111 (Supplementary Fig. S4). Notably, Tyr44 and Trp111 present potential hydrogen-bonding atoms towards the distal face of haem such that they could interact with a bound ligand. The side-chain N$^{\varepsilon1}$ atom of Trp111 lies some 2.9 Å from the modelled hydroxide molecule, while Tyr44 is at a distance of 3.5 Å but could interact more strongly with a bound oxygen molecule. We note that the ferric protein crystal may have become at least partially reduced in the X-ray beam, which could result in an iron(II)–oxy complex, although attempts to model a haem-bound oxygen were unsuccessful.

3.5. Structural and sequence comparisons

Analysis *versus* the PDB using *PDBeFold* (Krissinel & Henrick, 2004) shows that AHb3 has high structural homology to several bacterial truncated Hbs. A superposition with the Hbs from the actinobacterium *Thermobifida fusca* (PDB entry 2bmm; Bonamore *et al.*, 2005) and *Bacillus subtilis* (PDB entry 1ux8; Giangiacomo *et al.*, 2005) are given in Supplementary Fig. S5. Other than small variations in interhelical coil regions, the core tertiary structures are highly similar. Core r.m.s.d. values for superpositions by secondary-structure matching (Krissinel & Henrick, 2004) with the AHb3 structure in *Coot* (Emsley & Cowtan, 2004) were 1.19 Å for PDB entry 2bmm and 1.20 Å for PDB entry 1ux8. Sequence-alignment comparisons between AHb3 and class 3 proteins from *Hordeum vulgare*, *Medicago truncatula* and *Gossypium hirsutum* have been reported previously (Watts *et al.*, 2001). Since that publication, further class 3 plant globin sequences have been reported and alignments are presented in Supplementary Fig. S6. The sequence homology in the N-terminal region is extensive, suggesting similar quaternary structures, with the N-terminal region constituting the interface of the dimeric subunits. The C-terminal domain sequences show less sequence homology; however, one of the cysteine residues is

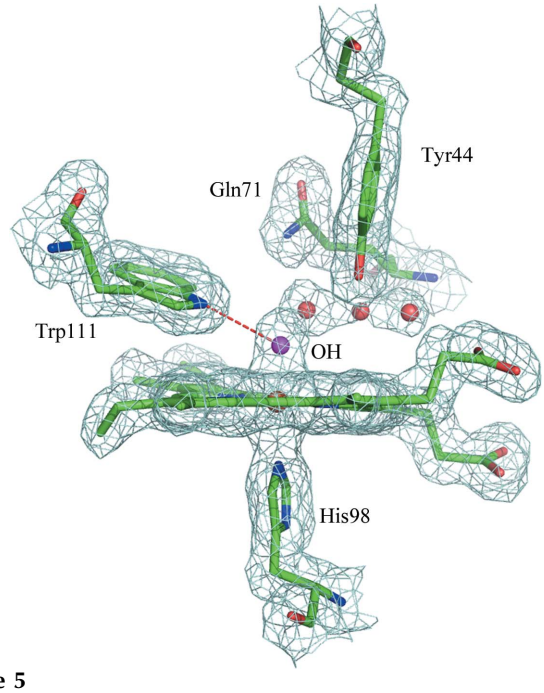

Figure 5
$2F_o - F_c$ electron-density map contoured at 1σ for the haem environment of AHb3. The Fe atom is six-coordinate with a proximal His98 ligand and a hydroxide modelled at the distal face. The bound hydroxide is positioned to form a hydrogen bond to Trp111 (red dashed line), while Gln71 and Tyr44 are too distant for such an interaction, although it is likely that they would interact with any bound dioxygen molecule in the oxy state of the protein. Several water molecules are present in the distal pocket with partial occupancies. This, together with weak $F_o - F_c$ difference density above the OH ligand, suggests that the crystal may have become partly reduced to the ferrous–oxy state in the X-ray beam.

largely conserved. This supports the hypothesis that the C-terminal strands from the two subunits are disulfide-linked *in vitro* in nonreducing cellular environments.

4. Discussion

The optical spectra of AHb3 were typical of pentacoordinate haemoglobin, with no evidence of endogenous hexacoordination from a distal protein ligand as was previously reported (Watts *et al.*, 2001). Erythrocyte haemoglobin can often form unstable low-spin hexacoordinate forms in the ferric oxidation state termed haemichromes. The formation of haemichromes is pH-dependent and temperature-dependent (Rifkind *et al.*, 1994; Sugawara *et al.*, 2003). Initial expression of AHb3 generated protein that was prone to precipitation and haemichrome formation. This was avoided in later expression protocols by preventing acidification of the protein following *E. coli* lysis. The spectra of ferric AHb3 at lower pH values show the 630 nm band that is representative of a water molecule associated with the haem iron in the distal coordination site. The acid–alkaline transition between the aqua and hydroxyl forms of ferric AHb3 at pH 7.17 lies at the lower end of the pH range characteristic of pentacoordinate haemoglobins. Typical examples include values of 8.99 for sperm whale myoglobin, 8.05 for human haemoglobin and 7.4 for *Chironomus* haemoglobin and *Aplysia* myoglobin (Antonini & Brunori, 1971; Brunori *et al.*, 1968; Scheler & Fischbach, 1958; Svistunenko *et al.*, 2007).

The function of cysteine residues in mammalian globins are often unambiguous, with human cytoglobin showing two intermolecular disulfide bonds forming the main link between the homodimeric subunits (Lechauve *et al.*, 2010). Neuroglobin shows an intramolecular disulfide bond integral to the tertiary protein structure and the affinity of the distal histidine for the haem iron (Hamdane *et al.*, 2005). It is noteworthy, therefore, that the disordered C-terminal region in AHb3 contains two cysteine residues at positions 163 and 165. With the structure showing that the final ordered residues before the disordered section of the protein are in close proximity in the two monomers, it is not unreasonable to propose that under appropriate cellular redox conditions the cysteines from each subunit will be sufficiently close to form disulfide bridges. Whether this linkage exists *in vivo*, or whether the cysteines form links to another molecule, remains unknown. Given that neither sequence nor structure predict any transmembrane or lipid anchor domains, such an association would be likely to occur *via* a protein–protein complex. If substantiated, this linkage could provide further clues to the physiological function of the protein.

A broad cavity on the proximal side of the haem is present, such that both haem propionates and the proximal His ligand are solvent-exposed. In contrast, the distal haem-pocket cavity is occupied by three ordered water molecules, as depicted in Fig. 5. The open structure of the haem pocket to external solvent may be related to the unusual concentration-independent binding kinetics for O_2 and CO following photodissociation as reported previously (Watts *et al.*, 2001). The authors speculated that this resulted from an unusually low rate of geminate recombination. However, the open nature of the haem pocket does not support this hypothesis and may be related to the transient hexacoordinate state of the protein. The cavity is similar to that observed in truncated Hbs and is not connected to bulk solvent *via* any observed waters. The lack of solvent-excluded cavities in AHb1 and bacterial truncated haemoglobins led to suggestions of NO dioxygenase activity (Daigle *et al.*, 2009; Hill, 2012; Thiel *et al.*, 2011). Although there are no large internal cavities suggestive of specific NO binding function, as have been observed in AHb1 and neuroglobin (Abbruzzetti *et al.*, 2009; Spyrakis *et al.*, 2013), an NO dioxygenase function cannot be ruled out from our structure. It is thus conceivable that the functional role of AHb3 could involve scavenging of NO generated by nitrite reductases.

BJR would like to thank the Royal Society (grant RG110485) for support. We thank Diamond Light Source for access to beamline I24 (East of England Macromolecular Crystallography BAG, MX7461) that contributed to the results presented here.

References

Abbruzzetti, S., Faggiano, S., Bruno, S., Spyrakis, F., Mozzarelli, A., Dewilde, S., Moens, L. & Viappiani, C. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 18984–18989.
Antonini, E. & Brunori, M. (1971). *Frontiers in Biology*, edited by A. Neuberger & E. L. Tatum, pp. 13–52. Amsterdam: North-Holland.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bonamore, A., Ilari, A., Giangiacomo, L., Bellelli, A., Morea, V. & Boffi, A. (2005). *FEBS J.* **272**, 4189–4201.
Bruno, S., Faggiano, S., Spyrakis, F., Mozzarelli, A., Abbruzzetti, S., Grandi, E., Viappiani, C., Feis, A., Mackowiak, S., Smulevich, G., Cacciatori, E. & Dominici, P. (2007). *J. Am. Chem. Soc.* **129**, 2880–2889.
Brunori, M., Amiconi, G., Antonin, E., Wyman, J., Zito, R. & Fanelli, A. R. (1968). *Biochim. Biophys. Acta*, **154**, 315–322.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Daigle, R., Rousseau, J. A., Guertin, M. & Lagüe, P. (2009). *Biophys. J.* **97**, 2967–2977.
Dalziel, K. & O'Brien, J. R. (1957). *Biochem. J.* **67**, 119–124.
Dordas, C., Rivoal, J. & Hill, R. D. (2003). *Ann. Bot.* **91**, 173–178.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Garrocho-Villegas, V., Gopalasubramaniam, S. K. & Arredondo-Peter, R. (2007). *Gene*, **398**, 78–85.
Giangiacomo, L., Ilari, A., Boffi, A., Morea, V. & Chiancone, E. (2005). *J. Biol. Chem.* **280**, 9192–9202.
Hamdane, D., Kiger, L., Dewilde, S., Uzan, J., Burmester, T., Hankeln, T., Moens, L. & Marden, M. C. (2005). *FEBS J.* **272**, 2076–2084.
Hill, R. D. (2012). *AoB Plants*, **2012**, pls004.
Hunt, P. W., Watts, R. A., Trevaskis, B., Llewelyn, D. J., Burnell, J., Dennis, E. S. & Peacock, W. J. (2001). *Plant Mol. Biol.* **47**, 677–692.
Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature (London)*, **185**, 422–427.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

Lechauve, C., Chauvierre, C., Dewilde, S., Moens, L., Green, B. N., Marden, M. C., Célier, C. & Kiger, L. (2010). *FEBS J.* **277**, 2696–2704.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Perazzolli, M., Dominici, P., Romero-Puertas, M. C., Zago, E., Zeier, J., Sonoda, M., Lamb, C. & Delledonne, M. (2004). *Plant Cell*, **16**, 2785–2794.

Perutz, M. F. (1960). *Brookhaven Symp. Biol.* **13**, 165–183.

Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L. & Bolognesi, M. (2000). *EMBO J.* **19**, 2424–2434.

Pesce, A., Dewilde, S., Nardini, M., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. & Bolognesi, M. (2004). *Micron*, **35**, 63–65.

Pesce, A., Nardini, M., Milani, M. & Bolognesi, M. (2007). *IUBMB Life*, **59**, 535–541.

Reeder, B. J., Cutruzzolà, F., Bigotti, M. G., Watmough, N. J. & Wilson, M. T. (2007). *IUBMB Life*, **59**, 477–489.

Rifkind, J. M., Abugo, O., Levy, A. & Heim, J. (1994). *Methods Enzymol.* **231**, 449–480.

Sanctis, D. de, Dewilde, S., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T. & Bolognesi, M. (2004). *J. Mol. Biol.* **336**, 917–927.

Scheler, W. & Fischbach, I. (1958). *Acta Biol. Med. Ger.* **1**, 194–210.

Smagghe, B. J., Hoy, J. A., Percifield, R., Kundu, S., Hargrove, M. S., Sarath, G., Hilbert, J. L., Watts, R. A., Dennis, E. S., Peacock, W. J., Dewilde, S., Moens, L., Blouin, G. C., Olson, J. S. & Appleby, C. A. (2009). *Biopolymers*, **91**, 1083–1096.

Spyrakis, F., Lucas, F., Bidon-Chanal, A., Viappiani, C., Guallar, V. & Luque, F. J. (2013). *Biochim. Biophys. Acta*, **1834**, 1957–1967.

Sugawara, Y., Kadono, E., Suzuki, A., Yukuta, Y., Shibasaki, Y., Nishimura, N., Kameyama, Y., Hirota, M., Ishida, C., Higuchi, N., Haramoto, K., Sakai, Y. & Soda, H. (2003). *Acta Physiol. Scand.* **179**, 49–59.

Svistunenko, D. A., Reeder, B. J., Wankasi, M. M., Silaghi-Dumitrescu, R. L., Cooper, C. E., Rinaldo, S., Cutruzzolà, F. & Wilson, M. T. (2007). *Dalton Trans.*, pp. 840–850.

Thiel, J., Rolletschek, H., Friedel, S., Lunn, J. E., Nguyen, T. H., Feil, R., Tschiersch, H., Müller, M. & Borisjuk, L. (2011). *BMC Plant Biol.* **11**, 48.

Watts, R. A., Hunt, P. W., Hvitved, A. N., Hargrove, M. S., Peacock, W. J. & Dennis, E. S. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10119–10124.

Yang, L.-X., Wang, R.-Y., Ren, F., Liu, J., Cheng, J. & Lu, Y.-T. (2005). *Plant Cell Physiol.* **46**, 1309–1316.