STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Structure of *Chlamydomonas reinhardtii* THB1, a group 1 truncated hemoglobin with a rare histidine–lysine heme ligation

Selena L. Rice,^{a,‡} Lauren E. Boucher,^{b,c,‡} Jamie L. Schlessman,^d Matthew R. Preimesberger,^a Jürgen Bosch^{b,c} and Juliette T. J. Lecomte^{a,*}

Received 2 March 2015

Accepted 7 April 2015

Edited by M. S. Weiss, Helmholtz-Zentrum Berlin für Materialien und Energie, Germany

‡ SLR and LEB contributed equally.

Keywords: 2/2 hemoglobin; distal ligand; nitric oxide dioxygenase; nitrogen metabolism; lysine coordination.

PDB reference: *Chlamydomonas reinhardtii* THB1, 4xdi

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

^aT. C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA, ^bDepartment of Biochemistry and Molecular Biology, Johns Hopkins University, Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA, ^cJohns Hopkins Malaria Research Institute, Johns Hopkins University, Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA, and ^dChemistry Department, US Naval Academy, 572 Holloway Road, Annapolis, MD 21402, USA. *Correspondence e-mail: lecomte_jtj@jhu.edu

THB1 is one of several group 1 truncated hemoglobins (TrHb1s) encoded in the genome of the unicellular green alga *Chlamydomonas reinhardtii*. THB1 expression is under the control of NIT2, the master regulator of nitrate assimilation, which also controls the expression of the only nitrate reductase in the cell, NIT1. *In vitro* and physiological evidence suggests that THB1 converts the nitric oxide generated by NIT1 into nitrate. To aid in the elucidation of the function and mechanism of THB1, the structure of the protein was solved in the ferric state. THB1 resembles other TrHb1s, but also exhibits distinct features associated with the coordination of the heme iron by a histidine (proximal) and a lysine (distal). The new structure illustrates the versatility of the TrHb1 fold, suggests factors that stabilize the axial ligation of a lysine, and highlights the difficulty of predicting the identity of the distal ligand, if any, in this group of proteins.

1. Introduction

THB1 is a cytosolic hemoglobin expressed by the green alga *Chlamydomonas reinhardtii*. Sequence analyses place it within the ‘truncated’ lineage of the hemoglobin superfamily, specifically in group 1 of this ancient branch (Wittenberg *et al.*, 2002). Group 1 truncated hemoglobins (TrHb1s) are present in a wide range of unicellular organisms (Vinogradov, Bailly *et al.*, 2013; Vinogradov *et al.*, 2011; Vinogradov, Tinajero-Trejo *et al.*, 2013). Little physiological information is available for these proteins, and their function is generally not known. THB1 is one of the few exceptions for which *in vivo* and *in vitro* data are being collected systematically. Its occurrence in a model alga makes THB1 an excellent candidate for further studies aimed at understanding the chemistry and molecular evolution of TrHb1s.

A BLAST search (Altschul *et al.*, 1990) using the THB1 sequence (UniProtKB accession No. A8JAR4) as a query retrieves two proteins with an identity higher than 50%. These are a hypothetical protein from *Volvox carteri* (D8THA8; 67% identity with a query cover of 91%) and THB2 from *C. reinhardtii* (A8JAR3; 66% identity with a query cover of 86%). Among the next closest 150 sequences (TrHbs with *E*-values lower than 10^{-20} and identities between 38 and 50%), *C. eugametos* LI637 (Q08753, 48% identity with a query cover of 83%) stands out as a eukaryotic protein with a TrHb1

domain (CtHb) for which some structural (Pesce *et al.*, 2000) and (bio)chemical (Couture *et al.*, 1994; Couture, Das *et al.*, 1999; Rice *et al.*, 2014) information is available. Thus, THB1 belongs to a well represented group of proteins but has a primary structure distant from its characterized relatives.

THB1 is linked to *C. reinhardtii* nitrogen metabolism (Johnson *et al.*, 2014). The *THB1* gene is regulated by NIT2, a transcription factor that controls the expression of the nitrate reductase gene (*NIT1* or *NIA1*) and other genes involved in nitrate assimilation. Accordingly, THB1 is absent from strains of *C. reinhardtii* with nonfunctional *NIT2*; furthermore, the protein is detectable only when the growth medium contains nitrate (Johnson *et al.*, 2014). Isolation of THB1 from *C. reinhardtii* cells confirms that the protein contains a *b* heme as its native cofactor (Johnson & Lecomte, 2014). *In vitro* studies show that ferrous recombinant THB1 binds exogenous ligands such as dioxygen, nitric oxide and carbon monoxide, and is capable of efficient nitric oxide dioxygenase (NOD) activity when provided with oxygen and a suitable reduction system (Johnson *et al.*, 2014). NOD activity is a plausible function of THB1 in the cytoplasm of *C. reinhardtii*, where it would eliminate nitrite-induced nitric oxide produced by NIT1 (Maia & Moura, 2015). Indirect and reversible inhibition of NIT1 by nitric oxide (Sanz-Luque *et al.*, 2013) suggests that NIT1 and THB1 constitute a cognate pair (Sanz-Luque *et al.*, 2015), with THB1 using NIT1 as its native reductase and preventing full reduction of the NIT1 catalytic site. This mechanism has been proposed for, and is supported by, a naturally occurring globin–nitrate reductase chimera (Stewart & Coyne, 2011).

In solution at neutral pH, recombinant THB1 is a 'hexa-coordinate' hemoglobin in both the ferric and ferrous states, *i.e.*, the heme iron has two axial ligands contributed by the protein. One of these is the proximal histidine, His77, a strictly conserved residue in functional hemoglobins. The other ligand, identified by amino-acid replacement and nuclear magnetic resonance (NMR) spectroscopy in the ferrous state, is Lys53 (Johnson *et al.*, 2014). The histidine–lysine heme ligation scheme has been observed only in a few cytochrome *c* proteins (Ferrer *et al.*, 1993; Rodrigues *et al.*, 2006). In THB1, hexacoordination may serve to facilitate electron transfer (Johnson & Lecomte, 2014) and attenuate the kinetics of exogenous ligand binding. In the present work, we describe the structure of THB1 with the distal lysine bound to the ferric heme and assess the structural factors associated with this rare form of heme coordination.

2. Materials and methods

2.1. Macromolecule production

Recombinant THB1 was expressed and purified as described previously (Johnson *et al.*, 2014). In brief, THB1 accumulated in inclusion bodies during overexpression in *Escherichia coli*. The inclusion bodies were solubilized with 8 M urea and the apoprotein was partially purified and refolded by gel-filtration chromatography. Hemin was added

Table 1

Macromolecule-production information.

Source organism	<i>C. reinhardtii</i>
DNA source	Synthesized gene (DNA 2.0)
Expression vector	pJExpress414
Expression host	<i>E. coli</i>
Complete amino-acid sequence of the construct produced	AADTAPADSLYSRMGGEAAVEKAVDVFYERIVAD-PQLAPFFANVDMKKQRKQVAFMTYVFGSGA-YEGRDLGASHRRLLIREQGMNHHFDLVAHLD-STLQELGVAQELKAEAMAIVASARPLIFGTGE-AGAAN
UniProt identifier	A8JAR4
Ligand	<i>b</i> heme (HEM)
Formula weight (Da)	15180.9

Table 2

Crystallization (crystals 1 and 2).

Method	Vapor diffusion, hanging drop
Plate type	VDX 24-well (Hampton Research)
Temperature (K)	298
Protein concentration (mg ml ⁻¹)	30
Buffer composition of protein solution	0.1 M glycine pH 9.5
Composition of reservoir solution	1.8 M ammonium sulfate, 20%(v/v) glycerol, 0.1 M glycine
Volume and ratio of drop	6 µl, 4:2 protein:mother liquor
Volume of reservoir (ml)	1

in excess to generate the ferric holoprotein, which was further purified by anion-exchange chromatography and exchanged into ~0.3 mM phosphate pH 7.5 before lyophilization. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Lyophilized THB1 was solubilized in 5 mM Tris pH 7.1 and made to a stock concentration of ~30 mg ml⁻¹. Solubilized THB1 was mixed with a buffer consisting of 0.1 M glycine pH 9.5 with varying concentrations of ammonium sulfate (1.6–2.1 M) and different amounts of glycerol [0–20%(v/v)] in a 4:2 ratio (protein:mother liquor) and incubated at room temperature. Red, hexagonal crystals of ferric THB1 first appeared after approximately two weeks in 0.1 M glycine pH 9.5, 1.8 M ammonium sulfate, 0, 10 or 20%(v/v) glycerol. Single crystals from the 20%(v/v) glycerol condition were flash-cooled in liquid nitrogen for data collection. Crystallization information is summarized in Table 2.

2.3. Data collection and processing

Data from two crystals were collected to solve the THB1 structure using multi-crystal SAD phasing. One data set was collected on a home source at Johns Hopkins University using an Agilent SuperNova sealed tube with a wavelength of 1.54 Å. A second high-resolution data set was collected on beamline X25 at the National Synchrotron Light Source (NSLS) and served as the native data set. Anomalous data were processed and scaled using the *CrysAlis^{Pro}* software from Agilent; the native data were processed using *XDSGUI* as a front end to *XDS/XSCALE/XDSConv* (Kabsch, 2010). SIRAS phasing using both data sets was performed using *HKL2MAP* (Pape & Schneider, 2004) as a front end to

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

	Native	Anomalous
Data-set name	JL01-4_5	xs0378a_max
Diffraction source	NSLS beamline X25	Sealed tube
Wavelength (Å)	1.1	1.54
Temperature (K)	100	110
Detector	PSI Pilatus 6M	Agilent ATLAS CCD
Crystal-to-detector distance (mm)	350	100
Rotation range per image (°)	0.3	0.3
Total rotation range (°)	105	242
Exposure time per image (s)	0.5	45
Space group	<i>P</i> ₆ 22	<i>P</i> ₆ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	144.51, 144.51, 79.91	144.62, 144.62, 78.80
α , β , γ (°)	90, 90, 120	90, 90, 120
Mosaicity (°)	0.097	0.75
Resolution range (Å)	47.30–1.89 (1.96–1.89)	50.00–3.00 (3.11–3.00)
Total No. of reflections	408679 (27314)	195664 (14343)
No. of unique reflections	39522 (3849)	36655 (3666)
Completeness (%)	99.9 (99.0)	99.3 (99.5)
Multiplicity	10.3 (7.1)	5.3 (3.9)
$\langle I/\sigma(I) \rangle$	13.8 (1.0)†	12.4 (3.1)
<i>R</i> _{meas}	0.08576	—
<i>CC</i> _{1/2}	0.999 (0.78)	—
<i>CC</i> *	1 (0.936)	—
Overall <i>B</i> factor from Wilson plot (Å ²)	39.20‡	—
SIRAS phasing		
Resolution (Å)		50–3.50
No. of heavy atoms in asymmetric unit		2
Heavy-atom type		Fe
(FOM) (initial/after <i>SHELXE</i>)		0.243/0.615

† Luo *et al.* (2014), Karplus & Diederichs (2012), Diederichs & Karplus (2013). ‡ No anomalies.

SHELXC/D/E (Sheldrick, 2010) to identify two Fe sites and determine initial phases and for subsequent autotracing. *SHELXE* placed 224 residues into electron density, yielding an almost complete trace of the model assuming the presence of two polypeptide chains within the asymmetric unit. Data-collection and processing statistics are presented in Table 3.

2.4. Structure solution and refinement

A single round of *Buccaneer* (Cowtan, 2006, 2008) in *CCP4* (Winn *et al.*, 2011) resulted in complete assignment of residues based on the recombinant THB1 sequence. *Coot* (Emsley *et al.*, 2010) and *phenix.refine* (Adams *et al.*, 2010) were used for subsequent rounds of model building and refinement, respectively. The heme ligands (HEM) were placed using *Coot*. Final refinement was performed using TLS groups, which were automatically determined *via* *PHENIX*. The quality of the structure was validated with *MolProbity* (Chen *et al.*, 2010). Modeling and refinement statistics are presented in Table 4.

3. Results and discussion

As purified from the *E. coli* host, the recombinant *C. reinhardtii* THB1 polypeptide has a cleaved initial methionine, is

Table 4

Structure solution and refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	47.30–1.89 (1.94–1.89)
Completeness (%)	99.9 (99.0)
σ Cutoff	$F > 1.36\sigma(F)$
No. of reflections, working set	39472 (2608)
No. of reflections, test set	1993 (125)
Final <i>R</i> _{cryst}	0.187 (0.378)
Final <i>R</i> _{free}	0.218 (0.386)
No. of non-H atoms	
Protein	1931
Ligand	86
Water	209
Total	2226
R.m.s. deviations	
Bonds (Å)	0.010
Angles (°)	1.12
Average <i>B</i> factors (Å ²)	
Protein	51.1
Ligand	39.7
Water	51.2
Ramachandran plot	
Most favored (%)	100
Allowed (%)	0

135 residues in length (Table 2) and lacks the N-terminal acetylation detected in the native protein (Johnson & Lecomte, 2014). Under the chosen conditions, THB1 crystallized in space group *P*₆22. The crystallographic asymmetric unit contained only two protein molecules, leading to a large solvent content (70.3%) and a high Matthews coefficient of 4.14 Å³ Da^{−1} as calculated in *CCP4* using *MATTHEWS_COEF* (Winn *et al.*, 2011).

The structure was solved using multi-crystal SAD phasing. Although several homologous structures with suitable sequence identity were available, it was not possible to find a molecular-replacement solution using either *MOLREP* (Vagin & Teplyakov, 2010) or *Phaser* (McCoy *et al.*, 2007). Part of this problem could be attributable to the hexagonal point group. It was initially unclear to which space group the crystal belonged, and the use of anomalous phasing helped to confirm the space group. Also, the high solvent content seemed to present a problem for both *MOLREP* and *Phaser*, as these programs would try to override the suggested two molecules in the asymmetric unit and try to place three molecules instead. When the programs did place two molecules only, they were incorrectly placed in the density.

3.1. Ferric THB1 structure: backbone and side chains

The structure was solved to a resolution of 1.9 Å (PDB entry 4xdi, Table 4). The two chains align with an r.m.s.d. of 0.53 Å over 119/123 C^α pairs (*UCSF Chimera*, interatomic distance threshold of 2 Å; Pettersen *et al.*, 2004), with the largest deviations located in the termini and loops. The electron density was insufficient to define the position of the N-terminal residues 2–4 in chain *A* (Ala-Ala-Asp) and 2–5 in chain *B* (Ala-Ala-Asp-Thr) and the C-terminal residues 132–136 in chain *A* (Ala-Gly-Ala-Ala-Asn) and 129–136 in chain *B* (Thr-Gly-Glu-Ala-Gly-Ala-Ala-Asn). Up to residue 12, the chains appear relatively unstructured and accessible to

solvent, which is consistent with the high affinity of recombinant THB1 for antibodies raised against the Ala2–Arg14 peptide (Johnson *et al.*, 2014). Chain *B* is composed of seven helical elements (Fig. 1*a*). These are labelled A–H (D is absent) as is customary for hemoglobins. Chain *A* contains the same secondary-structure elements, except that residues 13–15

are described as coil by *DSSP* (Kabsch & Sander, 1983). Overall, THB1 has the expected TrHb1 fold and common attributes, including a short or imperfectly formed A helix (Wenke *et al.*, 2014), a long EF loop, and N-terminally capped C 3_{10} -helix (Asp35–Gln37; Vuletich & Lecomte, 2006), E α -helix (Asp46–Lys49) and G α -helix (Asn87–His90).

Figure 1
(*a*) Ribbon diagram of THB1 (chain *B*). The helical segment identified by *DSSP* (Kabsch & Sander, 1983) spans residues 13–15 (A), 18–34 (B), 39–42 (C), 47–62 (E), 73–84 (F), 88–104 (G) and 109–127 (H). (*b*) Stereoview (wall-eyed) of the $2F_o - F_c$ electron-density map (contoured at the 1.0σ level; chain *A*) of the heme, its axial ligands (Lys53 and His77) and several residues within 4 Å of the heme group (C^α atoms depicted by balls). (*c*) $2F_o - F_c$ electron-density map (contoured at the 1.0σ level; chain *A*) of the heme and the residues forming well defined ionic interactions with the propionates (Lys49 and Arg52; C^α atoms depicted by balls). Distances are in Å. The electron-density maps in (*b*) and (*c*) were prepared with *PyMOL* (Schrödinger).

3.2. Ferric THB1 structure: heme axial ligands, geometry and environment

The conserved proximal histidine (His77; Figs. 1*a* and 1*b*) has its ring positioned such that steric clashes with the heme pyrrole N atoms are minimized. This is also observed in other TrHb1s and differs from the geometry in vertebrate oxygen-binding globins (Samuni *et al.*, 2004). A major distinguishing feature of ferric THB1 is the distal lysine at position 53 (E10 in the Perutz helical nomenclature; Perutz *et al.*, 1965). The side chain is well defined by the electron density (Fig. 1*b*) and extends almost parallel to the heme plane with *trans* (t) χ_1 , χ_2 and χ_3 dihedral angles and a *gauche*+ (p) χ_4 angle, which brings the N $^\zeta$ atom within 2.2 Å of the Fe atom and forms a C $^\epsilon$ —N $^\zeta$ —Fe angle of $\sim 118^\circ$. The heme environment on the distal side is largely hydrophobic (Fig. 1*b*), with no opportunity for Lys53 N $^\zeta$ H $_2$ to form hydrogen bonds. No water molecule was detected in the heme pocket.

Solvent accessibility to the set of nonpolar residues shown in Fig. 1*b*), other heme-cavity residues and the heme cofactor itself (except for the propionate head groups) is either very low or zero. Analysis of the heme geometry by normal coordinate structural decomposition (Jentzen *et al.*, 1998) reveals an out-of-plane displacement of ~ 1 Å, which is contributed almost equally by saddling and ruffling modes with opposite signs. This is in contrast to other TrHb1s, which tend to be predominantly ruffled (Wenke *et al.*, 2014). As depicted in Fig. 1*c*), the heme propionates are well defined. On the distal side, the amino group of Lys49 (E6) packs against the 6-propionate carboxylate and the guanidino group of Arg52 (E9) packs against the 7-propionate carboxylate. These hydrogen-bond and ionic interactions restrict entrance to the distal heme cavity.

3.3. Comparison with other TrHb1s

A search for the closest THB1 structural matches was performed with DALI (Holm & Rosenström, 2010). The top eight wild-type entries have Z-scores between 14.6 and 16.5 and are from six different TrHb1s (source organisms *Synechococcus* sp. PCC 7002, *Synechocystis* sp. PCC 6803, *C. eugametos*, *Paramecium caudatum*, *Tetrahymena pyriformis* and *Mycobacterium tuberculosis*; Supplementary Table S1). Some of these entries are discussed in the subsequent sections. Entries with Z-scores of between 10 and 15 correspond mostly to TrHbs (group 1, 2 and 3). Lower Z-scores (7–10) are dominated by proteins from the myoglobin-like lineage of the superfamily (Vinogradov, Tinajero-Trejo *et al.*, 2013), specifically proteins with the canonical 3/3 helical fold of the α subunit of vertebrate hemoglobin (no D helix). Non-hemoglobin proteins such as phycocyanin (Pastore & Lesk, 1990) have Z-scores of ~ 6 or below.

The structure of THB1 can be compared with those of the hexacoordinate TrHb1s (GlbNs) from the cyanobacteria *Synechococcus* (PDB entry 4max; Wenke *et al.*, 2014) and *Synechocystis* (PDB entry 1rtx; Hoy *et al.*, 2004). These two proteins use the shorter and more rigid His E10 as the distal ligand. Automatic alignment using UCSF Chimera (Pettersen

et al., 2004) matches large sections of the B, C, F and G helices and the beginning of the H helix with deviations below 2 Å (Supplementary Table S2, Supplementary Figs. S1*a* and S1*d*). The overlay with *Synechococcus* GlbN is presented in Fig. 2*a*). Within the globally optimized superposition, the heme planes are close to each other. Divergence of the H helices at the C-terminal end is associated with a cross-link between GlbN His117 and the heme 2-vinyl (Scott *et al.*, 2002). In THB1, the B and E helices are farther apart than in *Synechococcus* GlbN, whereas the E and F helices are closer together. This is likely to be related to the different constraints imposed by the coordinating lysine or histidine side chain.

Synechococcus and *Synechocystis* GlbN undergo a large conformational change upon exogenous ligand binding (Trent *et al.*, 2004; Wenke *et al.*, 2014). Although THB1 is hexacoordinate, some features of its heme environment resemble the structures of TrHb1s with exogenous ligand bound. One example is provided by Tyr29 (B10), an important residue because of its role in the stabilization of the oxy and other complexes (Pesce *et al.*, 2000, 2013; Nardini *et al.*, 2007; Igarashi *et al.*, 2011). In the hexacoordinate GlbNs, Tyr B10 points away from the heme cavity (Fig. 2*a*) and rotates in when the distal histidine is displaced by cyanide (Fig. 2*b*). In hexacoordinate THB1, the configuration of the B and E helices is such that Tyr B10 is positioned closer to its expected position in the cyanide complex. In fact, the structural superposition of THB1 with *Synechococcus* GlbN in the cyanide-bound state (PDB entry 4l2m; Wenke *et al.*, 2014; Fig. 2*b*; Supplementary Fig. S1*b*) is comparable to the superposition with the bis-histidine complex. In addition, each of these structural superpositions is arguably better than that of the bis-histidine and cyanide-bound states of *Synechococcus* GlbN with each other, if one ignores the post-translationally modified H helix (Supplementary Table S2, Supplementary Fig. S1*c*). Superpositions of THB1 with *Synechocystis* GlbN and CtrHb in the cyanide-bound state [PDB entries 1s69 (Trent *et al.*, 2004) and 1dly (Pesce *et al.*, 2000), respectively] and *M. tuberculosis* HbN in the oxy state (PDB entry 1idr; Milani *et al.*, 2001) are also similar (Supplementary Table S2, Supplementary Figs. S1*e*, S1*f*, S1*g* and S2). We conclude that THB1 has structural traits of both the His–His and the His–diatomic complexes. The structure illustrates that the loops of TrHb1s, in particular the EF loop (Dellarole *et al.*, 2013; Pond *et al.*, 2012), allow the helices that they connect to adopt a broad range of relative orientations.

A few TrHb1 structures that have a distal exogenous ligand exhibit tunnels providing access to the heme cavity (Milani *et al.*, 2001; Daigle *et al.*, 2009; Pesce *et al.*, 2013; Wenke *et al.*, 2014). Analysis of the structure of THB1 with coordinated lysine using the program MOLEonline v.2.0 (Berka *et al.*, 2012) reveals no such tunnel. However, both chains present a narrow pore with one end located between the GH turn and the opposing B helix and the other end between the G and H helices (Supplementary Figs. S3*a* and S3*b*). In chain B Phe91 (G5) terminates the pore, whereas a constriction at the level of G9 blocks access to the heme in chain A. The characteristics of the pore are a minimum radius of 0.8 Å, a length of 26 Å and a

volume of $\sim 180 \text{ \AA}^3$ in chain A. In chain B, the same parameters are $1.0 \text{ \AA}$, $25 \text{ \AA}$ and $\sim 230 \text{ \AA}^3$, respectively. For comparison, the corresponding tunnel in CtrHb (Milani *et al.*, 2001) is shown in Supplementary Fig. S3(c) (parameters of $1.35 \text{ \AA}$, $30 \text{ \AA}$ and $\sim 370 \text{ \AA}^3$, respectively). The pore detected in THB1 is expected to widen once an exogenous ligand binds to the iron on the distal side.

3.4. Other instances of lysine coordination

The lysine coordination observed in THB1 requires that the protein withstand the energetic cost associated with the deprotonation of the amino group at physiological pH and the immobilization of an otherwise flexible side chain. Few examples of lysine ligation have been characterized structurally thus far: lysine with an opposing axial histidine as in THB1 for ferric M100K cytochrome c_{550} (PDB entry 2bh5;

Worrall *et al.*, 2005; Ubbink *et al.*, 1994) and heme 4 of NrfH in the cytochrome *c* nitrite reductase NrfHA complex from *Desulfovibrio vulgaris* (PDB entry 2j7a; Rodrigues *et al.*, 2006) and lysine with an opposing water molecule in NrfA of the same complex (Rodrigues *et al.*, 2006).

The coordinating lysine side chains have no common geometry; they can be parallel (THB1), at an angle (M100K cytochrome c_{550}) or perpendicular (nitrite reductase) to the heme plane (Supplementary Fig. S4). There is no trend in the dihedral angles or the secondary structure of the segment bearing the lysine (Supplementary Table S3). It is interesting that the previously available examples of lysine ligation all involve *c* hemes. THB1 demonstrates that covalent attachment of the heme group to the polypeptide is not a requirement. The amino group of the lysine may or may not form a hydrogen bond to a distal residue or a water molecule (Supplementary Fig. S4, Supplementary Table S3). Never-

Figure 2

Overlay of THB1 with *Synechococcus* GlnB (a) in the bis-histidine state (PDB entry 4max; Wenke *et al.*, 2014) and (b) in the cyanide-bound state (PDB entry 4l2m; Wenke *et al.*, 2014). THB1 is in sand colour. Also represented are Tyr B10 and the axial His46 (E10) and His70 (F8) of GlnB along with the cross-linked His117 (H16) and the axial Lys53 (E10) and His77 (F8) of THB1. Linkers between helices are omitted for clarity.

theless, two common features emerge, in that the coordinating lysine is generally not directly accessible to bulk solvent and alternative de-coordinated states are expected to require structural rearrangement.

3.5. Structural predictions

Many TrHb1s have a lysine at position E10, but not all Lys E10 residues coordinate the iron. Two examples are ferric *M. tuberculosis* HbN (Couture, Yeh *et al.*, 1999) and *P. caudatum* HbN (Pesce *et al.*, 2000). In both instances a water molecule takes the place of Lys E10. So far, TrHb1 structures with exogenous ligand bound resemble each other closely (Supplementary Table S2); thus, using any of these structures as a template in programs such as *SWISS-MODEL* (Arnold *et al.*, 2006) to represent a TrHb1 of unknown structure is likely to yield acceptable results. However, a 'hexacoordinate' template (be it with axial His E10 or Lys E10) may also generate plausible structures. Such is the case for two other *C. reinhardtii* TrHb1s, THB2 (Sanz-Luque *et al.*, 2015) and THB8 (Hemschemeier *et al.*, 2013), which have a lysine at E10. Interestingly, an acceptable hexacoordinate model of *C. eugametos* TrHb1 (CtrHb) can also be prepared from THB1, and NMR data of the alkaline ferrous state are consistent with a coordinated lysine as in THB1 (Supplementary Fig. S5). It is unclear whether lysine ligation persists in the ferric state of CtrHb or whether the protein uses Tyr B10 as a distal ligand as previously proposed (Das *et al.*, 1999). In conclusion, the structure of THB1 illustrates that the TrHb1 fold can accommodate different distal heme ligands and emphasizes the difficulty in predicting whether a protein side chain coordinates the iron in the absence of a strong ligand. Structural determination by NMR or X-ray crystallography remains a necessity to characterize the heme environment and rationalize the chemical properties of these proteins.

4. Related literature

The following reference is cited in the Supporting Information for this article: Giangiacomo *et al.* (2005).

Acknowledgements

The authors thank Maxime Siegler for the collection of the anomalous data set and Eric Johnson for useful comments. The work was supported by funding from The Bloomberg Family Foundation (JB), the NIH Chemistry–Biology Interface Training Grant (T32 GM080189; LEB), the NIH Cellular and Molecular Biology Training Grant (T32 GM007231; SLR), a Predoctoral Fellowship from the Johns Hopkins Malaria Research Institute (LEB) and the National Science Foundation (MCB 1330488; JTTJL).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). *Bioinformatics*, **22**, 195–201.
 Berka, K., Hanák, O., Sehnal, D., Banáš, P., Navrátilová, V., Jaiswal, D., Ionescu, C.-M., Svobodová Vařeková, R., Koča, J. & Otyepka, M. (2012). *Nucleic Acids Res.* **40**, W222–W227.
 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.* **D66**, 12–21.
 Couture, M., Chamberland, H., St-Pierre, B., Lafontaine, J. & Guertin, M. (1994). *Mol. Gen. Genet.* **243**, 185–197.
 Couture, M., Das, T. K., Lee, H. C., Peisach, J., Rousseau, D. L., Wittenberg, B. A., Wittenberg, J. B. & Guertin, M. (1999). *J. Biol. Chem.* **274**, 6898–6910.
 Couture, M., Yeh, S.-R., Wittenberg, B. A., Wittenberg, J. B., Ouellet, Y., Rousseau, D. L. & Guertin, M. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 11223–11228.
 Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
 Cowtan, K. (2008). *Acta Cryst.* **D64**, 83–89.
 Daigle, R., Rousseau, J. A., Guertin, M. & Lagüe, P. (2009). *Biophys. J.* **97**, 2967–2977.
 Das, T. K., Couture, M., Lee, H. C., Peisach, J., Rousseau, D. L., Wittenberg, B. A., Wittenberg, J. B. & Guertin, M. (1999). *Biochemistry*, **38**, 15360–15368.
 Dellarole, M., Roumestand, C., Royer, C. & Lecomte, J. T. J. (2013). *Biochim. Biophys. Acta*, **1834**, 1910–1922.
 Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
 Ferrer, J. C., Guillemette, J. G., Bogumil, R., Inglis, S. C., Smith, M. & Mauk, A. G. (1993). *J. Am. Chem. Soc.* **115**, 7507–7508.
 Giangiacomo, L., Ilari, A., Boffi, A., Morea, V. & Chiancone, E. (2005). *J. Biol. Chem.* **280**, 9192–9202.
 Hemschemeier, A., Düner, M., Casero, D., Merchant, S. S., Winkler, M. & Happe, T. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 10854–10859.
 Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
 Hoy, J. A., Kundu, S., Trent, J. T. III, Ramaswamy, S. & Hargrove, M. S. (2004). *J. Biol. Chem.* **279**, 16535–16542.
 Igarashi, J., Kobayashi, K. & Matsuoka, A. (2011). *J. Biol. Inorg. Chem.* **16**, 599–609.
 Jentzen, W., Ma, J.-G. & Shelnutt, J. A. (1998). *Biophys. J.* **74**, 753–763.
 Johnson, E. A. & Lecomte, J. T. J. (2014). *FI1000Research*, **3**, 294.
 Johnson, E. A., Rice, S. L., Preimesberger, M. R., Nye, D. B., Gilevicius, L., Wenke, B. B., Brown, J. M., Witman, G. B. & Lecomte, J. T. J. (2014). *Biochemistry*, **53**, 4573–4589.
 Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
 Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
 Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
 Luo, Z., Rajashankar, K. & Dauter, Z. (2014). *Acta Cryst.* **D70**, 253–260.
 Maia, L. B. & Moura, J. J. G. (2015). *J. Biol. Inorg. Chem.* **20**, 403–433.
 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.
 Milani, M., Pesce, A., Ouellet, Y., Ascenzi, P., Guertin, M. & Bolognesi, M. (2001). *EMBO J.* **20**, 3902–3909.
 Nardini, M., Pesce, A., Milani, M. & Bolognesi, M. (2007). *Gene*, **398**, 2–11.
 Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
 Pastore, A. & Lesk, A. M. (1990). *Proteins*, **8**, 133–155.
 Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–678.
 Pesce, A., Bolognesi, M. & Nardini, M. (2013). *Adv. Microb. Physiol.* **63**, 49–78.
 Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L. & Bolognesi, M. (2000). *EMBO J.* **19**, 2424–2434.

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Pond, M. P., Majumdar, A. & Lecomte, J. T. J. (2012). *Biochemistry*, **51**, 5733–5747.
- Rice, S. L., Preimesberger, M. R., Johnson, E. A. & Lecomte, J. T. J. (2014). *J. Inorg. Biochem.* **141**, 198–207.
- Rodrigues, M. L., Oliveira, T. F., Pereira, I. A. & Archer, M. (2006). *EMBO J.* **25**, 5951–5960.
- Samuni, U., Ouellet, Y., Guertin, M., Friedman, J. M. & Yeh, S.-R. (2004). *J. Am. Chem. Soc.* **126**, 2682–2683.
- Sanz-Luque, E., Ocaña-Calahorra, F., de Montaigu, A., Chamizo-Ampudia, A., Llamas, Á., Galván, A. & Fernández, E. (2015). *Plant J.* **81**, 467–479.
- Sanz-Luque, E., Ocaña-Calahorra, F., Llamas, A., Galvan, A. & Fernandez, E. (2013). *J. Exp. Bot.* **64**, 3373–3383.
- Scott, N. L., Falzone, C. J., Vuletich, D. A., Zhao, J., Bryant, D. A. & Lecomte, J. T. J. (2002). *Biochemistry*, **41**, 6902–6910.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Stewart, J. J. & Coyne, K. J. (2011). *Plant Mol. Biol.* **77**, 565–575.
- Trent, J. T. III, Kundu, S., Hoy, J. A. & Hargrove, M. S. (2004). *J. Mol. Biol.* **341**, 1097–1108.
- Ubbink, M., Campos, A. P., Teixeira, M., Hunt, N. I., Hill, H. A. & Canters, G. W. (1994). *Biochemistry*, **33**, 10051–10059.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Vinogradov, S. N., Bailly, X., Smith, D. R., Tinajero-Trejo, M., Poole, R. K. & Hoogewijs, D. (2013). *Adv. Microb. Physiol.* **63**, 391–446.
- Vinogradov, S. N., Fernández, I., Hoogewijs, D. & Arredondo-Peter, R. (2011). *Mol. Plant.* **4**, 42–58.
- Vinogradov, S. N., Tinajero-Trejo, M., Poole, R. K. & Hoogewijs, D. (2013). *Biochim. Biophys. Acta*, **1834**, 1789–1800.
- Vuletich, D. A. & Lecomte, J. T. J. (2006). *J. Mol. Evol.* **62**, 196–210.
- Wenke, B. B., Lecomte, J. T. J., Héroux, A. & Schlessman, J. L. (2014). *Proteins*, **82**, 528–534.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A. & Guertin, M. (2002). *J. Biol. Chem.* **277**, 871–874.
- Worrall, J. A., Van Roon, A. M., Ubbink, M. & Canters, G. W. (2005). *FEBS J.* **272**, 2441–2455.