Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

Research Article

Crystal structure of a hemerythrin-like protein from *Mycobacterium kansasii* and homology model of the orthologous Rv2633c protein of *M. tuberculosis*

Zhongxin Ma[1], Jan Abendroth[2], Garry W. Buchko[3,4], Kyle H. Rohde[1] and Victor L. Davidson[1]

[1]Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL 32827, U.S.A.; [2]UCB Pharma, Bainbridge Island, WA 98110, U.S.A.; [3]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, U.S.A.; [4]School of Molecular Biosciences, Washington State University, Pullman, WA 99164, U.S.A.

Correspondence: Victor L. Davidson (victor.davidson@ucf.edu)

Pathogenic and opportunistic mycobacteria have a distinct class of non-heme di-iron hemerythrin-like proteins (HLPs). The first to be isolated was the Rv2633c protein, which plays a role in infection by *Mycobacterium tuberculosis* (*Mtb*), but could not be crystallized. This work presents the first crystal structure of an ortholog of Rv2633c, the mycobacterial HLP from *Mycobacterium kansasii* (*Mka*). This structure differs from those of hemerythrins and other known HLPs. It consists of five α-helices, whereas all other HLP domains have four. In contrast with other HLPs, the HLP domain is not fused to an additional protein domain. The residues ligating and surrounding the di-iron site are also unique among HLPs. Notably, a tyrosine occupies the position normally held by one of the histidine ligands in hemerythrin. This structure was used to construct a homology model of Rv2633c. The structure of five α-helices is conserved and the di-iron site ligands are identical in Rv2633c. Two residues near the ends of helices in the *Mka* HLP structure are replaced with prolines in the Rv2633c model. This may account for structural perturbations that decrease the solubility of Rv2633c relative to *Mka* HLP. Clusters of residues that differ in charge or polarity between Rv2633c and *Mka* HLP that point outward from the helical core could reflect a specificity for potential differential interactions with other protein partners *in vivo*, which are related to function. The *Mka* HLP exhibited weaker catalase activity than Rv2633c. Evidence was obtained for the interaction of *Mka* HLP irons with nitric oxide.

Introduction

The gene *Rv2633c* in *Mycobacterium tuberculosis* (*Mtb*) is of interest as it is up-regulated within minutes of phagocytosis of *Mtb* by macrophages during infection [1,2]. This rapid up-regulation, which occurs at least partly in response to phagosomal acidification, is mediated by two well-characterized regulators required for *Mtb* virulence, PhoPR and WhiB3 [3–5]. The potential importance of the Rv2633c protein during infection was highlighted by an *in vivo* TraSH screen which found that transposon insertions in Rv2633c significantly attenuated *Mtb* in a mouse infection model [6].

The primary amino acid sequence of the Rv2633c protein revealed that it is a hemerythrin-like protein (HLP) with an HHE (histidine–histidine–glutamic acid) cation-binding domain [7]. Hemerythrins do not contain heme but use non-heme irons for oxygen transport or storage [8]. These proteins typically are comprised of identical subunits of four α-helices, with each subunit housing a di-iron site with a bridging oxygen. The proteins typically possess six or eight of these subunits [8,9], although monomeric hemerythrins have been identified in the bacterium *Methylococcus capsulatus* [10] and the peanut worm *Themiste hennahi* [11]. The recombinant Rv2633c protein that was expressed in *Escherichia coli* and purified [7] was a homodimer that contained two iron ions per

Received: 11 November 2019
Revised: 6 January 2020
Accepted: 8 January 2020

Accepted Manuscript online: 8 January 2020
Version of Record published: 31 January 2020

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

subunit. However, it did not exhibit changes in its absorbance spectrum on the addition of reductant and O_2 that were characteristic of hemerythrin. It did exhibit catalase activity, which was unexpected for a hemerythrin and highly unusual for a non-heme di-iron active site.

Other HLPs have been characterized that also have a core structure comprised of four α-helices that house a di-iron site. In contrast with the true hemerythrins, the helical domain is connected to an additional protein domain. The nature of the additional domain confers distinct biological functions, which often relate to signal transduction via methyl-accepting chemotaxis, diguanylate cyclase, or adenlyl cyclase domains [12]. The ligation pattern of the di-iron site also varies with function among these HLPs, both in the number and identity of the amino acid residues that provide ligands for the two iron ions [9]. Other HLPs include an oxygen-sensing protein from *Desulfovibrio vulgaris* [13], an iron sensing protein in humans [14], and the YtfE protein from *E. coli* [15]. YtfE, which also possesses a bound flavin, catalyzes the repair of Fe–S centers in other proteins and exhibits nitric oxide reductase activity.

Structural studies of Rv2633c have been hampered by the instability of the protein in solution and its tendency to precipitate. As such, an ortholog approach was employed to obtain structural information on this protein [16]. A sequence and phylogenetic analysis revealed that this protein is a member of a subset of HLPs that are exclusive to pathogenic and opportunistic mycobacteria [7]. In the present study, a gene that encodes an HLP homologous to Rv2633c was cloned from *Mycobacterium kansasii* (*Mka*). The amino acid sequence of the 161 residue *Mka* HLP, as deduced from the gene sequence, is 78% identical and 86% similar to that of Rv2633c (from this point referred to as *Mtb* HLP), with no gaps in the sequence. The orthologous *Mka* HLP was expressed in *E. coli*, and milligram quantities of the relatively stable protein were obtained. The solution properties of the *Mka* HLP, its crystallization and crystal structure are presented. As suggested by previous phylogenetic analysis [7], the structure reveals that this mycobacterial HLP is structurally distinct from other classes of hemerythrin-like proteins. It differs in the number of helices in the core structure that contains the di-iron site. It lacks an additional protein domain and is monomeric. The ligation pattern and environment of the two iron ions is also different from those of true hemerythrins and previously described HLPs.

The *Mka* HLP structure was used as a reference to construct a homology model of the *Mtb* HLP. This model revealed that the residues that provide ligands for the two irons, as well as interacting residues in the second coordination sphere, are strictly conserved in the two proteins. The model also identified the locations in the structure of each of the amino acids in the sequences that differ between the *Mka* and *Mtb* HLPs. Specific residue changes were identified that provide the structural basis for differences in solution properties of the *Mka* and *Mtb* HLPs. The new *Mka* HLP structure and *Mtb* HLP model significantly expand our view of the structures and potential functions of HLPs.

Materials and methods

Cloning of the recombinant *M. kansasii* HLP

For heterologous expression of the *Mka* HLP, the MKAN_RS24640 gene was cloned from *Mka* chromosomal DNA (derived from an *Mka* clinical isolate kindly provided by the National Jewish Hospital). This was inserted into a pET15a vector, which adds a hexa-histidine tag at the N-terminus to facilitate purification. Briefly, the pET15a-*Mka* HLP recombinant expression plasmid was created by FastCloning [17] the MKAN_RS24640 coding sequence into the pET15a vector (Novagen). The MKAN_RS24640 gene was PCR amplified from chromosomal DNA purified from *Mka* using the primers MykaK.20209_pET15a-FC-F (5′-*CAGCAGCCATCATCATCATCAC*GTGAACGCATACGAAGTGCTCAAAG-3′) and MykaK.20209_pET15a-FC-R (5′-*GGCTTTGTTAGCAGCCGGATCC*TCACAGGCTTCGGACCAGCG-3′). The pET vector portion was PCR amplified using the primers pET15a_2633c_FC-F (5′-GGATCCGGCTGCTAACAAAGCC-3′) and pET15_Rev_noThromb (5′-GTGATGATGATGATGATGGCTGCTG-3′). Italicized letters in primer sequences indicate 5′ extensions that are complementary to the vector primers, which are required for the restriction enzyme- and ligase-free Fastcloning method. The DNA fragments were combined, digested with DpnI to eliminate parental plasmid DNA, and transformed into *E. coli* 10-beta (New England Biolabs). The transformed *E. coli* containing the plasmid pET15a-*Mka* HLP was plated on LB agar plates containing ampicillin (100 μg/ml). Positive clones were identified by PCR screening and confirmed by sequencing. DNA from the pET15a-MyKA2633c positive clone was then transformed into Rosetta 2 (DE3) *E. coli* (MilliporeSigma Novagen) for expression and was plated on LB agar plates containing the antibiotics chloramphenicol (34 μg/ml) and ampicillin (100 μg/ml). Rosetta 2 positive clones were identified by PCR screening and confirmed by sequencing.

Expression and purification of the *Mka* protein

For the protein used for the iron-bound *Mka* HLP structure, cells were grown in LB medium containing ampicillin and chloramphenicol, and supplemented with 7 mg/l Fe(III), which had been filter sterilized and added after autoclaving. Cells were grown at 30°C to an OD of ∼0.8 and induced with 1 mM IPTG, at same temperature, for 4 h. Cells were harvested by centrifugation and the cell pellet was frozen. A lysozyme/osmotic shock treatment was used to break the cells. The cells were suspended (0.2 g/ml) in buffer containing 20 mM Tris–HCl pH 8.0 plus 500 mM NaCl, with 10 mg lysozyme per g cells, PMSF (100 mM, 1 µl/ml), DNAse and a few drops of 1 M $MgCl_2$. The suspension was shaken at 30°C for 30 min. Then an equal volume of H_2O was added and the suspension was shaken for additional 30 min. The suspension was centrifuged to remove the solids and the supernatant was applied to a Nickel NTA affinity resin in the suspension buffer and eluted with an increasing gradient of imidazole.

For the protein used for the apo *Mka* HLP structure, cells were grown in autoinduction minimal medium (500 ml) supplemented with ampicillin (100 µg/ml), and chloramphenicol (34 µg/ml) [18]. Upon reaching an OD_{600} of ∼0.8, the culture temperature was reduced from 37 to 25°C and left to shake overnight. Cells were harvested mid-morning by mild centrifugation and frozen at −80°C. After generating a cell extract from the thawed cell pellet using a lysozyme/sonication method, the protein was purified using two chromatography steps: immobilized metal ion affinity with 20 ml Ni-Agarose 6 FastFlow column and Superdex75 gel-filtration with a HiLoad 26/60 column (GE Healthcare, Piscataway, NJ). The second chromatography step, in addition to removing minor impurities from the metal ion affinity step, exchanged *Mka* HLP into the buffer used for crystallization screening: 100 mM NaCl, 20 mM Tris–HCl, 1.0 mM dithiothreitol, pH 7.1 [19].

Protein analysis

The purity of the protein was assessed by SDS–PAGE. The native mass of the protein was determined by size exclusion chromatography using a HiPrep 16/60 Sephacryl S-300 HR column on an AKTA PURE FPLC system (GE Healthcare). The iron content of the protein was determined by a colorimetric assay [20] using ferrozine (Sigma). Results were calibrated to a standard curve of solutions of known concentrations of ferrous iron. Solution studies were performed at 37°C in 50 mM potassium phosphate buffer, pH 7.5. The catalase reaction rate was monitoring as described previously [7] by the decrease in absorbance at 240 nm, which corresponds to the concentration of H_2O_2 ($\varepsilon_{240} = 43.6\ M^{-1}\ cm^{-1}$). The protein was reduced by the addition of sodium dithionite. Nitric acid was added to solutions using the NO donor, diethylamine NONOate sodium (1,1-Diethyl-2-hydroxy-2-nitroso-hydrazine sodium).

Crystallography

Crystallization conditions were searched for using apo *Mka* HLP at 2 mg/ml using crystallization screens JCSG+, JCSG-Top96 (RigakuReagents), MCSG-1 (Anatrace), Morpheus, LMB, and BCS (Molecular Dimensions) using 0.4 + 0.4 µl drops and XJR crystallization trays (RigakuReagents). Crystals were found within a few days in the MCSG-1 screen condition F3: 200 mM ammonium citrate, 25% PEG 3350. Crystals for the high-resolution apo *Mka* HLP data set were vitrified with 20% ethylene glycol by plunging them into liquid nitrogen. Data were collected at the APS beamline 21-ID-F.

To phase the structure, a crystal from the same drop was incubated for 15 s in a solution containing 10% v/v of 5 M sodium iodide in ethylene glycol and 90% v/v reservoir. This was followed by another 15 s incubation in a solution containing 20% v/v of 5 M sodium iodide in ethylene glycol and 80% v/v reservoir, and vitrified by plunging them in liquid nitrogen. A 180° data set was collected on an in-house Rigaku F-RE$^+$ SuperBright X-ray generator with CuKα radiation (wavelength = 1.5418 Å). Diffraction data were reduced with XDS [21]. For the anomalous iodide data set; Friedel mates were kept separate. For the iodide data set, eight iodide sites were located with phenix.hyss [22]. The iodide sites were refined with Phaser_EP [23] yielding a phase set with an overall FOM of 0.377. Phases were improved with PARROT [24] to a FOM of 0.672. An initial model was built with ARPwARP [25]. The structure was then completed with iterative cycles of refinement in phenix.refine [22] and real space model building in Coot [26]. The quality of the model was assessed with tools built into Coot and phenix.refine, such as Molprobity [27].

Crystallization conditions for the iron-bound *Mka* HLP (22 mg/ml) were searched for using the MCSG-1 screen (Anatrace) and an optimization screen around the crystallization condition for the apoprotein. Crystals that were isomorphous to the apoprotein crystals grew from MCSG-1 condition (3.5 M sodium formate,

Biochemical Journal (2020) 477 567–581
https://doi.org/10.1042/BCJ20190827

100 mM sodium acetate/HCl pH 4.6). These crystals were vitrified directly in liquid nitrogen without adding additional cryoprotectant for in-house data collection. Initial data sets showed a low occupancy of iron that decreased during crystal growth. Optimizing the crystallization process, the highest iron-occupancy was observed in slightly modified conditions (3.0 M sodium formate, 100 mM sodium acetate/HCl pH 4.9) and crystals were harvested 15 h after tray set up. Data reduction and structure refinement used the same techniques as for the apo *Mka* HLP structure.

Homology modeling

A homology model of Rv2633c, the *Mtb* HLP, was generated using the Swiss-Model Homology modeling online software (https://swissmodel.expasy.org) [28] using the structure of the *Mka* HLP as a template.

Results

Solution properties of the *Mka* HLP

The purification of the *Mka* HLP yielded ∼5 mg of protein per liter of bacterial cell culture. The protein migrated on SDS–PAGE with a mass of ∼18 kDa. When subjected to size exclusion chromatography, the *Mka* HLP eluted as a single peak with a retention time expected for a protein with a mass of ∼18 kDa, as determined by comparison with molecular mass standards. The mass of the protein that is predicted from the sequence is 18 376 Da. Thus, this protein in solution is a monomer. The iron content of the protein as determined by the ferrozine assay [20] was 2.1 ± 0.1 irons per molecule from three different preparations. The iron was relatively tightly bound. After incubation for 1 h with 1 mM EDTA, 1.7 ± 0.1 irons per molecule were present. An apo *Mk*a HLP, which lacked iron, was also expressed and purified. When subjected to size exclusion chromatography, the apo *Mka* HLP eluted as two bands with retention times expected for a monomer and a dimer. The apoprotein had no visible absorbance spectrum, consistent with the lack of iron.

The absorbance spectrum of the purified *Mka* HLP exhibited a broad peak centered at ∼350 nm that overlaps the protein absorbance centered at 280 nm. This is attributed to the ferric non-heme irons (Figure 1A). In addition to the 350 nm absorbance, the oxidized protein exhibited a weak broad absorbance centered at ∼500 nm. On reduction by the addition of sodium dithionite under anaerobic conditions, this visible absorbance was bleached (Figure 1B). On re-exposure to air, the absorbance returned. When the reduced *Mka* HLP was exposed to NO under anaerobic conditions, a similar result was obtained with the formation of the peak at ∼350 nm, although the intensity of the peak was greater than in the as-isolated protein (Figure 1B). This indicates that NO is able to interact with the di-iron site. This unexpected result suggests a potential role for this protein in adaptation to nitrosative stress that the bacterium experiences during infection. The addition of NO to the as-isolated sample in aerobic conditions resulted in an enhancement of the visible absorbance peak (Figure 1C). As NO is expected to react with Fe(II), this suggests that some of the iron in the air-oxidized sample may be reduced. In addition to the reactivity towards NO, the *Mka* HLP also exhibited catalase activity which was weak compared with that previously reported for *Mtb* HLP [7]. *Mka* HLP exhibited a k_{cat} of 34 s^{-1} and K_m of 36 mM, compared with a k_{cat} of 1475 s^{-1} and K_m of 10 mM for *Mtb* HLP.

Structures of the *Mka* HLP

Initial attempts to crystallize the *Mka* HLP yielded crystals that lacked iron. As such, an initial structure was solved of the iron-free apo *Mka* HLP. It was subsequently observed that crystals of the iron-bound *Mka* HLP could be formed, but that they rapidly lost iron. This realization allowed an approach to obtain the structure of the holoenzyme as well. Despite the protein stock solution exhibiting a red color and the absorbance spectrum indicating the presence of ferric iron, there was no evidence for iron in the initial electron density maps. On closer inspection, it was noticed that the crystals formed within less than 1 day. A crystal that was harvested within 15 h of the crystallization set up resulted in an iron-bound structure with an iron-occupancy of 75% for Fe1 and 59% for Fe2. Data sets collected at longer times (two or more days) after crystallization set up resulted in the absence of electron density for iron, despite the fact that the irons are tightly bound to the protein in solution. The crystal structures of the apo and iron-bound of *Mka* HLPs were each refined to 1.75 Å (Table 1). A search using the EMBO-EBI PDBeFOLD server indicated that no homologous structure exists in the PDB.

The structure of the *Mka* HLP monomer consists of five α-helices. The portion of the structure that contains the four helices that provide ligands for the di-iron center is very similar to that of hemerythrin, which lacks the fifth helix (Figure 2A). The active site of the *Mka* HLP contains two iron ions (Figure 2B). Each iron is

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

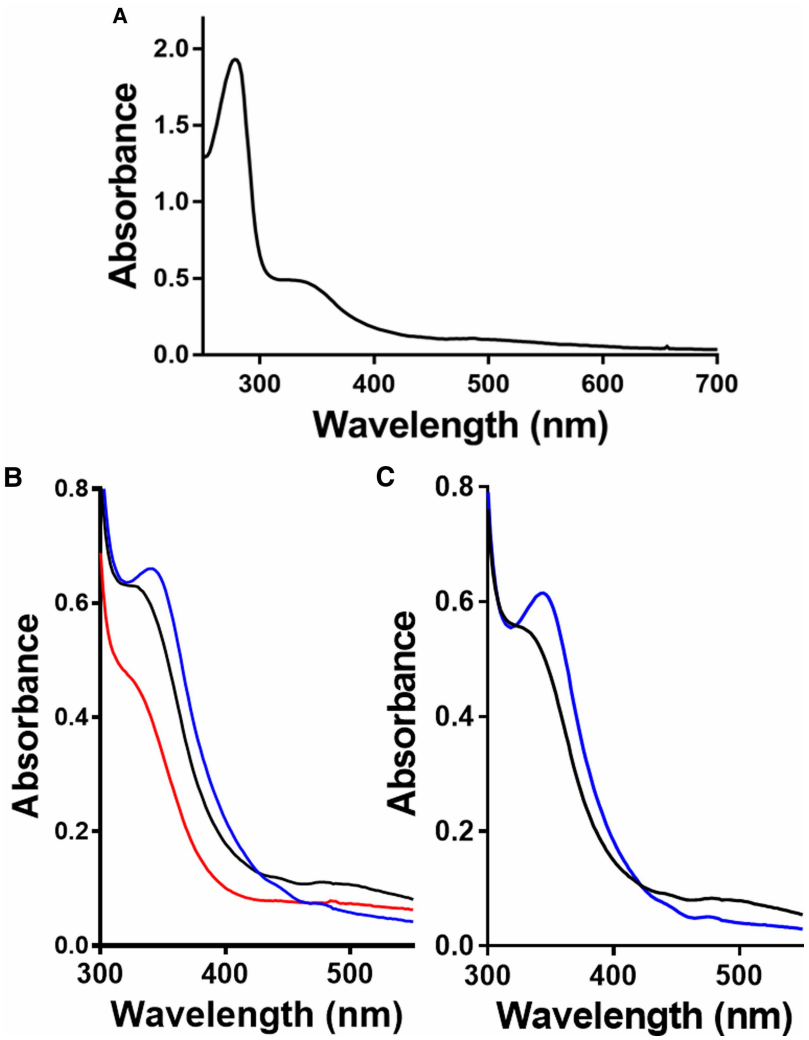

Figure 1. Absorption spectra of *Mka* HLP.

(**A**) Spectrum of the as isolated purified *Mka* HLP. (**B**) Amplification of the relevant region of the spectrum of *Mka* HLP as isolated (black), after anaerobic reduction by dithionite (red) and after subsequent anaerobic addition of NO (blue). (**C**) Amplification of the relevant region of the spectrum of *Mka* HLP as isolated (black) and after addition of NO in aerobic conditions (blue).

ligated by two His residues. In addition, two Glu residues each provide an oxygen ligand to each iron (Figure 2C). In all hemerythrins and HLP structures to date, a bridging oxygen links the two irons; this oxygen could be a water molecule or a hydroxide ion. Oxygen is also present in the structure of *Mka* HLP that interacts with both irons. This bridging oxygen/hydroxide is asymmetrically bound, with 1.71 Å distance to Fe1 and 2.48 Å distance to Fe2. This suggests the possibility of differing oxidation states of the two irons. A Tyr side chain is also present in the site with the phenyl oxygen within 3 Å of an iron, and thus could be considered to be a ligand. The distances between the irons and ligating residues are given in Table 2. Each iron has a different coordination geometry. Fe1 has five coordination partners (His11-NE2, His45-NE2, Glu49-OE2, Glu109-OE1, and the bridging oxygen), which form a square pyramid around Fe1 with the bridging oxygen at the tip. Fe2 has six coordination partners (Glu49-OE1, Tyr54-OH, His71-NE2, His105-NE2, Glu109-OE2, and the bridging oxygen) in a distorted octahedral geometry.

The overall structures of the holo *Mka* HLP and apo *Mka* HLP are essentially the same. However, there are differences in the positions of some residues that provide ligands to the irons in the di-iron site (Figure 3). In the apoprotein structure, the side chains of His11 and His71 are present in two conformations with

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

Table 1 Data collection, refinement, and structure quality of the *Mka* HLP

Parameter	Iodide derivative	Apo	Iron bound
PDB ID		6U3L	6Q09
Space group	$P3_121$	$P3_121$	$P3_121$
Unit-cell parameters			
a (Å)	52.21	52.05	52.49
b (Å)	52.21	52.05	52.49
c (Å)	104.28	104.70	104.42
$\alpha = \beta$, γ (°)	90, 120	90, 120	90, 120
Matthews coefficient ($Å^3\,Da^{-1}$)		2.10	2.13
Solvent content (%)		41.45	42.28
Resolution range (Å)	50–2.00 (2.05–2.00)*	50–1.75 (1.80–1.75)	50–1.75 (1.80–1.75)
Mean $I/\sigma(I)$	13.63 (2.65)	24.94 (2.18)	26.56 (3.66)
No. of observed unique reflections	20 880 (1176)†	17 189 (1264)	17 427 (1235)
Completeness (%)	97.3 (75.8)	100 (99.9)	99.9 (99.6)
Multiplicity	5.4 (3.8)	6.9 (4.9)	10.7 (11.0)
R^2_{merge}	0.087 (0.541)	0.041 (0.639)	0.045 (0.621)
CC 1/2	99.8 (73.0)		
Phasing			
Figure of merit (before density modification)	0.38		
Number of iodide sites	8		
Refinement			
No. of used reflections		17 176	17 335
R_{work} (%)[3]		0.178	0.183
R_{free} (%)[4]		0.218	0.225
Mean B factor ($Å^2$)		36.2	40.3
RMSD bonds (Å)		0.008	0.009
RMSD angles (°)		0.851	0.927
Model validation			
MolProbity			
Clash score, all atoms		4.67	3.02
Ramachandran favored (%)		99.4	99.4
Ramachandran outliers (%)		0	0
Rotamer favored (%)		99.3	92.3
Rotamer outliers (%)		0	2
MolProbity score		1.24	1.20

*Value in parenthesis are statistics for the highest resolution shell;
†The phasing iodide data show unmerged reflections, whereas the refined data sets show merged reflections.

approximately the same occupancy. For each His, one of the conformations in the apoprotein is the same as in the iron-bound protein. In the other, the side chain is pointing in a different direction. Thus, the presence of iron locks these two His side chains into one conformation, while in the apoprotein structure the energetic difference between the two conformations is small enough that both occur with similar occupancy/probability. In addition, the side chain of Glu109 in the apoprotein is flipped 90° relative to the position in the holoprotein where it interacts with the irons.

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

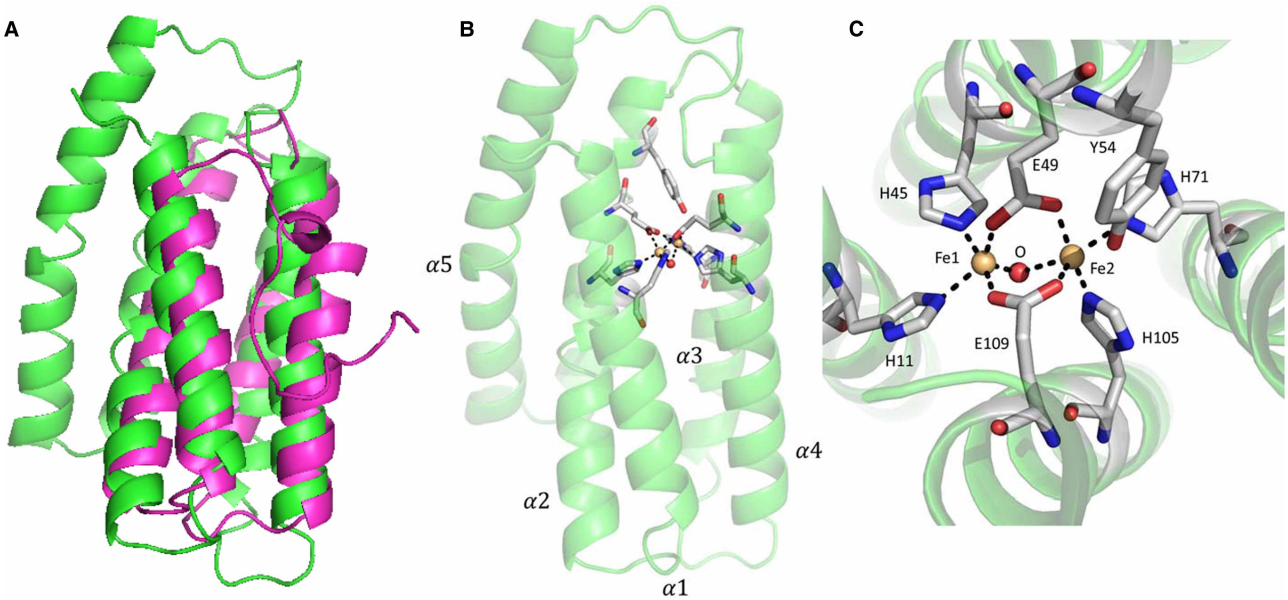

Figure 2. The structure of *Mka* HLP.

(**A**) Superimposition of the structures of *Mka* HLP in green and a monomeric hemerythrin from *Themiste hennahi* (PDB entry 1A7D) [11] in purple.
(**B**) The overall structure of *Mka* HLP. The five α-helices are labeled and the components of the di-iron site are drawn in stick representation.
(**C**) Expansion of the *Mka* HLP di-iron site with the ligand side chains labeled.

Homology model of Rv2633c, the *Mtb* HLP

The *Mka* HLP was cloned, expressed, and structurally characterized as part of an effort to obtain a soluble and stable ortholog of the Rv2633c protein from *Mtb*, which could not be crystallized. The sequences of the two proteins are 78% identical and 86% similar (Figure 4). Thus, one would intuitively expect the structures to be very similar. However, there must be key distinctions to account for differences in the properties of the two proteins in solution. The *Mtb* HLP [7] is a dimer and is poorly soluble, as it readily precipitates from solution. In contrast, the *Mka* HLP is a monomer that is relatively stable in solution. To gain insight into the details of the structure of the *Mtb* HLP, a homology model was constructed using its amino acid sequence and the *Mka* HLP structure as a template. Homology modeling was performed with the Swiss-Model Homology modeling online software [28]. The structure predicted for the *Mtb* HLP is very similar to the crystal structure of *Mka* HLP, as illustrated by the superimposition in Figure 5. The QMEAN Z-score [29] provides an estimate of the quality of the structural model. Scores around zero indicate good agreement between the model structure and experimental structures of similar size and scores below −4.0 indicate models with low quality. The QMEAN Z-score for the model of the *Mtb* HLP was −0.68, which means the modeled structure has a very high probability of being close to the actual structure. With the homology model, it was possible to identify the positions in the structures of each of the amino acid differences between the two proteins in order to provide potential

Table 2 The distances between irons with ligands in the active site

Bond	Distance (Å)	Bond	Distance (Å)
Fe1–Fe2	3.19	Fe1–OE1 (E109)	1.96
Fe1–O (bridge)	1.71	Fe2–NE2 (H71)	2.37
Fe2–O (bridge)	2.48	Fe2–NE2 (H105)	2.05
Fe1–NE2 (H11)	2.15	Fe2–OE1 (E49)	2.05
Fe1–NE2 (H45)	1.92	Fe2–OE2 (E109)	1.98
Fe1–OE2 (E49)	2.12	Fe2–OH (Y54)	2.96

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

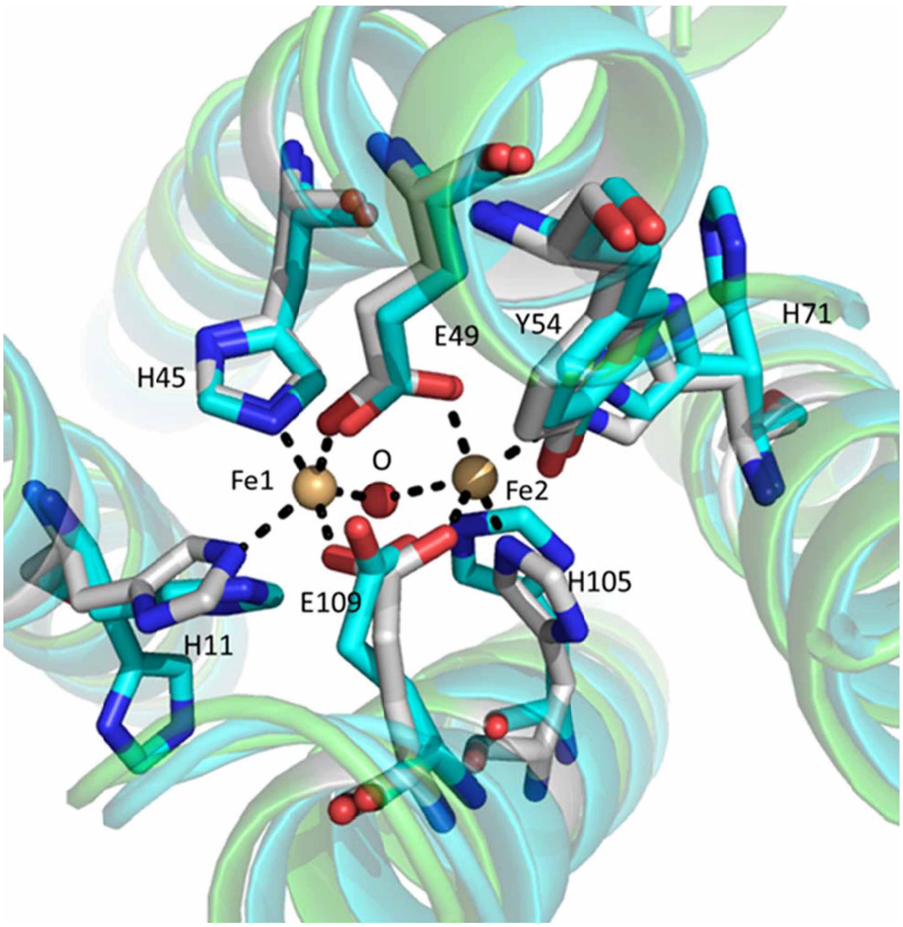

Figure 3. Superimposition of the active sites in the crystal structures of iron-bound *Mka* HPL (carbons in gray) and the apoprotein (carbons in blue).

The two iron atoms and the bridging oxygen, shown as spheres, are from the holoprotein structure. For the apoprotein structure, side chains of H11 and H71 were each modeled in two conformers with approximately equal occupancy.

```
Mka      1  MNAYEVLKEHHVVIKGLGRKISEAPVNSEERHALFDELLIELDIHFRIEDDLYYPALSAA  60
            MNAY+VLK HH V+KGLGRK+ EAPVNSEERH LFDE+LIELDIHFRIEDDLYYPALSAA
Mtb      1  MNAYDVLKRHHTVLKGLGRKVGEAPVNSEERHVLFDEMLIELDIHFRIEDDLYYPALSAA  60

Mka     61  TKLIAVAHAEHRQVIDQLSVLLRTPQSEPGYEDEWNSFKTVLEAHADEEERDMIPAPPEV  120
             K I   HAEHRQV+DQL+ LLRTPQ  PGYE+EWN F+TVLEAHAD EERDMIPAP  V
Mtb     61  GKPITGTHAEHRQVVDQLATLLRTPQRAPGYEEEWNVFRTVLEAHADVEERDMIPAPTPV  120

Mka    121  KITDAELEELGEKMAARMEQYRGSALYKLRTKGRAALVRSL  161
             ITDAELEELG+KMAAR+EQ RGS LY LRTKG+A L++++
Mtb    121  HITDAELEELGDKMAARIEQLRGSPLYTLRTKGKADLLKAI  161
```

Figure 4. Sequence alignment of the HLPs from *Mka* and *Mtb*.

The residues in the HHE cation binding domains of each protein are colored red. Identical residues are listed in the line between the protein sequences and similar residues are indicated with a plus (+) sign.

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

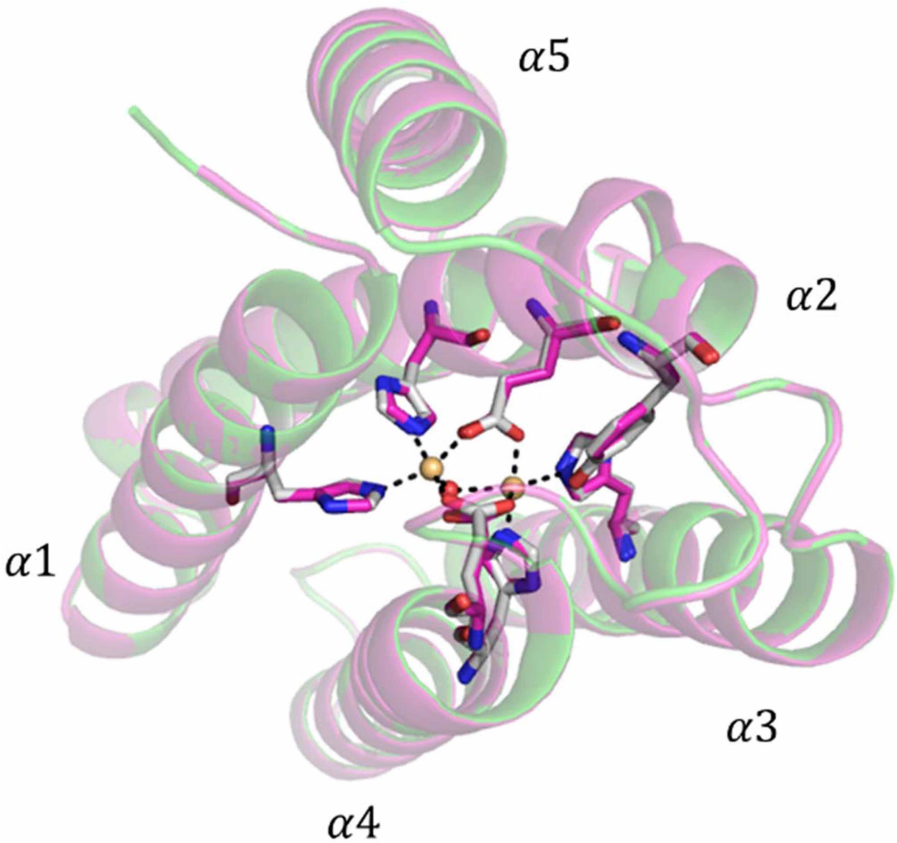

Figure 5. Superimposition of the iron-bound *Mka* HLP structure (green) and *Mtb* HLP homology model (pink) with the irons atoms from the *Mka* HLP structure shown.The side chains of the residues involved in chelating the iron atoms are shown in stick representation and are in essentially identical positions in the crystal structure and the homology model.

insight into the different physical properties of the two proteins. These differences are not randomly distributed throughout the structure but clustered in certain regions (Figure 6). One of the most notable distinctions is that Pro residues are present in the *Mtb* HLP in positions that reside in α-helices in the *Mka* HLP structure.

The residues in helices α1 and α2 are relatively conserved. Helix α1 (residues 3–23) has variations in six of 21 residues. For the most part, the changes result in residues with similar properties going from *Mka* to *Mtb*. Three of them point toward solvent (Glu5 to Asp5, Glu9 to Arg9, and Val12 to Thr12). The other three point toward the inside of the protein (Ile14 to Leu14, Ile21 to Val21, and Ser22 to Gly22). The reversal of charge of residue 9 is the most notable difference. His11, which provides an iron ligand, is on this helix. Helix α2 (residues 29–57) is predominantly located in the interior of the protein with little surface exposure. It is highly conserved with only one change to a similar residue, Ala33 to Val33. Three of the residues that provide iron ligands reside on this helix, His45, Glu49, and Tyr54.

Helix α3 (residues 61–83) exhibits the largest variation between the *Mka* and *Mtb* proteins, with changes in eight of the 23 residues. A significant change is Leu63 to Pro in the *Mtb* HLP. This residue is near the beginning of the helix and likely alters the conformation or shortens the helix in a way not reflected in the homology model. The stretch of residues 61–67 is the least conserved portion of the overall protein sequences (see Figure 4). This helix also contains the His71 iron ligand. The loop that connects helix α3 and helix α4 also contains two non-conservative amino acid differences, Ser87 to Arg87 and Glu88 to Ala88. Helix α4 (residues 90–113) exhibits only four changes out of the 24 residues. Two of the residues providing iron ligands are present on this helix, His105 and Glu109. The changes result in similar residues except for Glu108 to Val108. However, the loop connecting helix α4 and helix α5 have three non-conservative changes, Pro118 to Thr118, Glu119 to Pro 119, and Arg121 to His121 from *Mka* to *Mtb*.

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

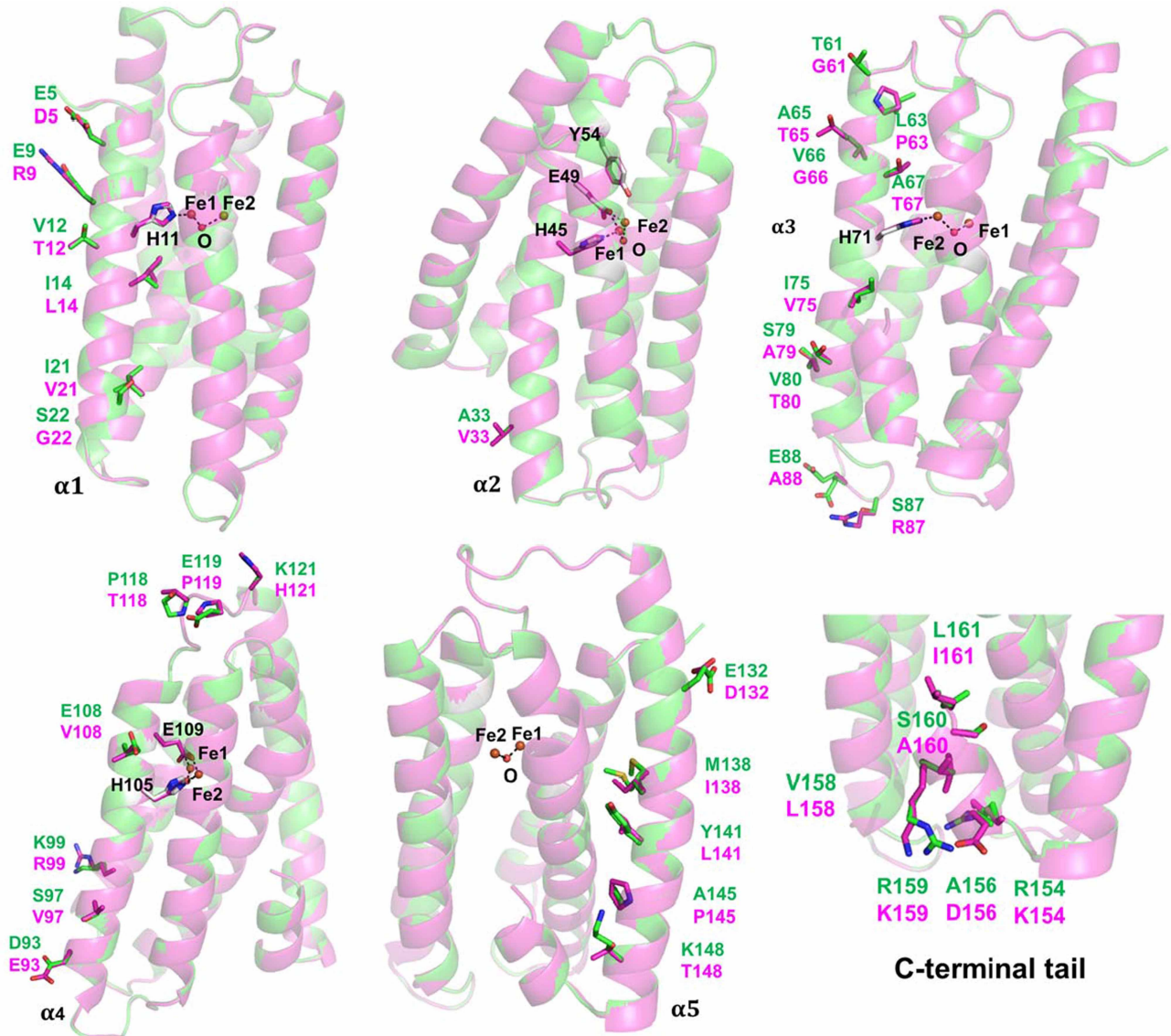

Figure 6. Superimposition of the *Mka* HLP structure (green) and *Mtb* HLP homology model (pink).

The six panels highlight differences in the amino acid sequences and the orientation of each of the five α-helices and the C-terminal tail. At positions of amino acid differences, the side chains in the *Mtb* HLP model are illustrated in stick representations.

Helix α5 (residues 124–152) is only present in the mycobacterial HLPs. There are no iron ligands on this helix. There are five changes in these 29 residues, the potentially most significant being Ala145 to Pro145 in the *Mtb* HLP. Again, this suggests that this helix in the *Mtb* HLP may be altered relative to what is seen in the homology model. In contrast with the other helices, where many of the residues that were changed pointed to the outside of the protein, four of the five differing residues in this helix point towards the inside. One additional structural feature that exhibits variation is the eight residues that comprise the C-terminal tail. Six of these residues vary from *Mka* to *Mtb*, although the features of the side chains are similar. This segment exhibits some α-helical character but is very short. Furthermore, this segment does not extend from the body of the protein but resides in a pocket formed by helices α2, α3, and α5.

With respect to the residues that provide ligands for the two iron ions, the di-iron site in the *Mtb* HLP aligns perfectly with the *Mka* HLP (Figure 7). In fact, in addition to the strict conservation of position of the primary ligands in the first coordination sphere, amino acids in a second coordination sphere that interact with

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

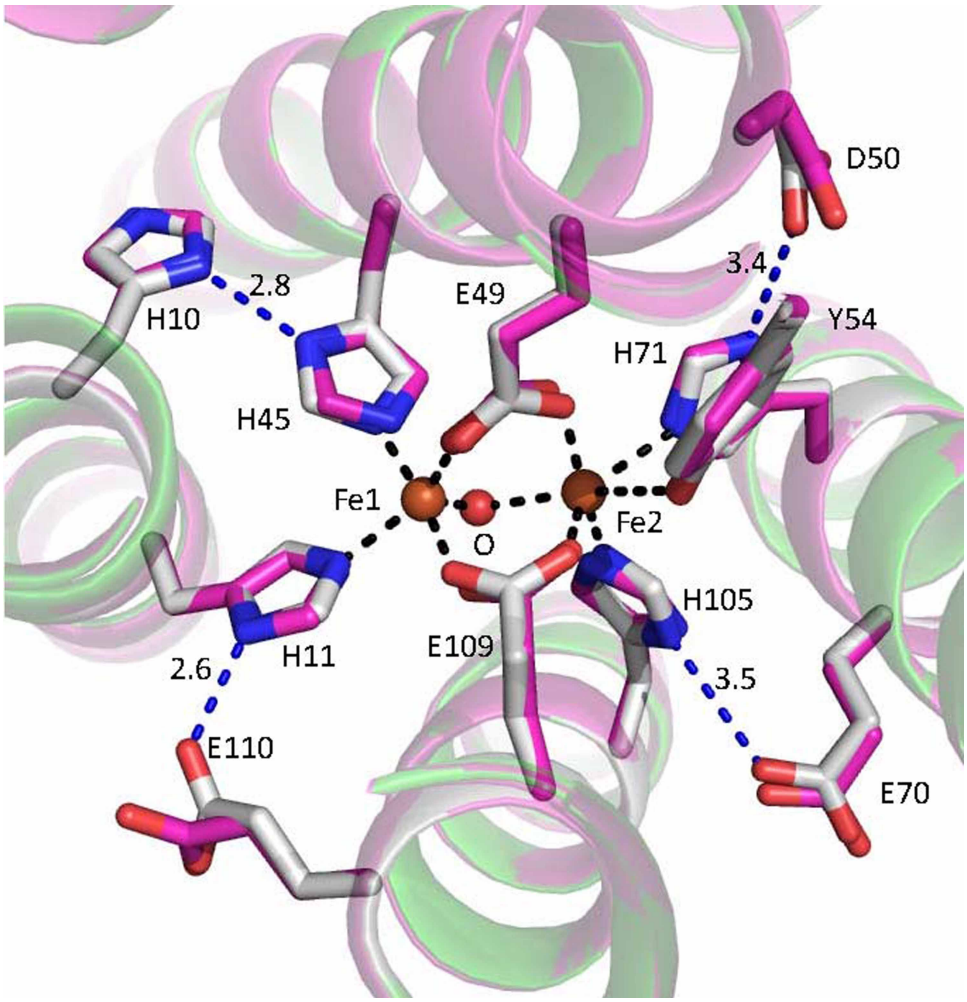

Figure 7. Superimposition of the first and second coordination spheres of the *Mtb* HLP and *Mka* HLP.
The residue numbers are the same for both proteins. The carbons on the *Mka* HLP residues are gray and those on the HLP *Mtb* residues are pink. The nearest points of contact between side-chain atoms bound to the iron and side chains in the second coordination sphere are illustrated with dashed lines with the interatomic distance given in Ångstroms.

the primary iron ligands are also conserved through either ionic or hydrogen bonding interactions within 3.5 Å. Of the four His ligands of the irons, two interact with Asp side chains, one with a Glu side chain and one with another His side chain. This strict conservation of ligands and residues that surround the di-iron site was not necessarily expected. The four α-helix core of hemerythrin and other classes of HLPs have similar overall structures, but variation in the residues in the di-iron sites (Figure 8). In contrast, the environment of the di-iron site is retained in these different mycobacteria.

Discussion

There are notable differences of the structures of these mycobacterial HLPs relative to true hemerythrins and all other known HLPs. They possess a fifth α-helix in addition to the four-helix core present in the other proteins. The ligation pattern of the iron ions in the di-iron site is also distinct from the other proteins (Figure 8). True hemerythrins, which function in reversible O_2 binding exhibit a di-iron site with a bridging oxygen, in which one iron has ligands provided by three His residues and the other iron is co-ordinated by two His residues. In addition, a Glu and an Asp each provide one oxygen ligand to each iron. In contrast, mycobacterial HLPs have a Tyr residue instead of a ligating His residue, which could interact with the iron. Hemerythrins and HLPs are typically comprised of identical subunits of four α-helices that each house a di-iron site. The

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

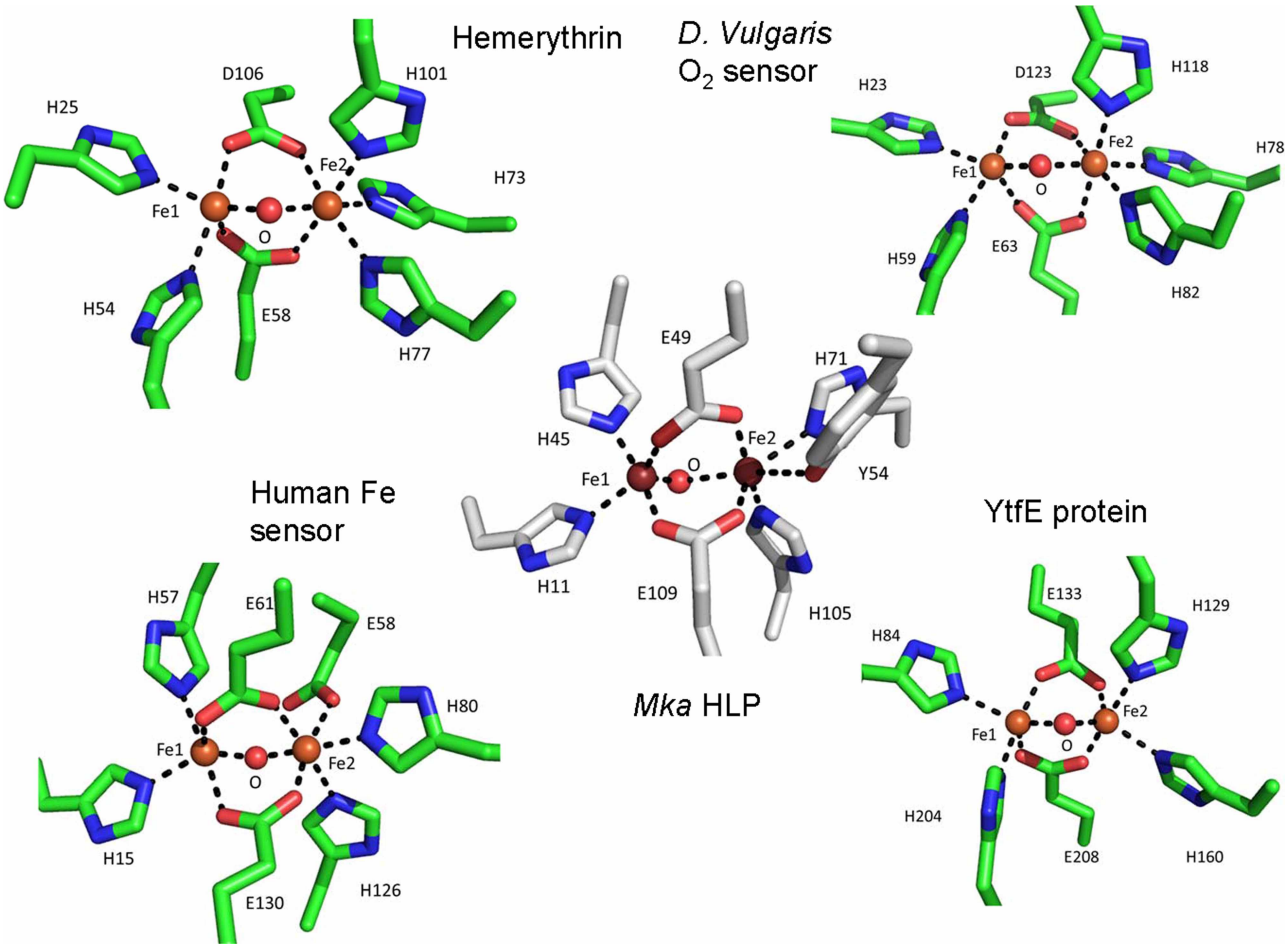

Figure 8. Di-iron centers of the *Mka* HLP and other known HLPs.

The PDB entries for the structures that are shown are Mka HLP (6Q09), *Themiste dyscrita* hemerythrin (1HMD) [33], *Desulfovibrio vulgaris* O_2 sensor (2AWC) [13], *Homo sapiens* iron sensor (3V5X) [14], and *E. coil* YtfE (5FNN) [15]. Carbons are colored gray in the *Mka* HLP and green in the other proteins.

native proteins are comprised of six or eight of these subunits [8,9] although a bacterial hemerythrin has been identified that is a monomer [10]. The other HLPs with different iron ligation patterns and different biological functions possess additional domains on each protein that are required for their respective biological functions. An oxygen-sensing protein has been described from *D. vulgaris*, which has a di-iron site with the same ligation pattern as true hemerythrins [13]. An HLP domain has been described in a protein found in humans that functions as an iron sensor. It has a ligation pattern in which each iron has two ligands provided by His residues and three Glu residues providing additional oxygen ligands [14]. The HLP domain which bears the most similarity to the *Mka* HLP is the YtfE protein from *E. coli* which also has ligands from two His residues and two Glu residues [15]. However, as stated earlier the mycobacterial HLPs have a Tyr residue in the position normally occupied by the additional His ligand. Furthermore, YtfE possesses a bound flavin cofactor, catalyzes the repair of Fe–S centers in other proteins, and has nitric oxide reductase activity. The lack of additional domains of known function further distinguishes the mycobacterial HLPs from the other HLPs. Considering this observation, one cannot rule out the possibility that other mycobacterial proteins could form a complex with a mycobacterial HLP to confer a biological function, similar to the role of the additional protein domains in the other HLPs. For example, as stated above, YtfE requires the fused FAD-containing domain for NO reductase activity. As the *Mka* HLP exhibits reactivity towards NO, it is possible that it could function as an NO reductase given an appropriate electron donor or redox protein partner.

While the di-iron sites in the *Mka* and *Mtb* HLPs are identical, outside of the di-iron site the homology model of the *Mtb* HLP reveals notable differences on certain residues. Proline residues are present near the ends of helices α3 and α5 in the *Mtb* protein. It is possible that in the true *Mtb* structure the helical character of those two regions is perturbed by these Pro residues, and this results in larger disordered segments connecting helices α2 and α3, and helix α5 with the C-terminal tail. This could be related to the instability and poor solubility of the isolated *Mtb* HLP relative to the *Mka* HLP. Helices α3 and α4 also exhibit several differences in residues on the protein surface, which point outwards towards the solvent from each helix. This could be related to the fact that *Mtb* HLP is a dimer in solution and iron-bound *Mka* HLP is a monomer. It also suggests the possibility that these residues could be involved in protein–protein interactions with other proteins *in vivo* that differ in the two hosts. The possibility that such unrecognized protein–protein interactions could be important for function should not be ruled out, as these mycobacterial HLPs are the only ones that do not possess an additional protein domain that is required for biological function. The *Mka* HLP crystal structure and *Mtb* HLP homology model provide a framework for future site-directed mutagenesis studies to test such hypotheses. The new structure and model also significantly expand our view of the diverse structures and functions of HLPs.

It should also be noted that *Mycobacterium kansasii* does not cause tuberculosis. However, *Mka* is an opportunistic pathogen that causes pulmonary disease indistinguishable from tuberculosis. *Mka* is phylogenetically the most closely related non-tuberculous mycobacterial species to *Mtb* [30]. Although the incidence is low in healthy individuals, *Mka* is the second most common non-tuberculous mycobacterial species in HIV/AIDS patients [31]. Detailed studies of *Mka* pathogenesis and host interactions have been limited. However, like *Mtb*, intracellular survival within macrophages is an important virulence mechanism of *Mka* [32]. Thus, the host-derived cues and stresses and the corresponding survival strategies are expected to be similar for *Mka* and *Mtb*. While the low solubility of the *Mtb* HLP in solution complicates structure-function studies in solution, the results of this study will inform *in vitro* studies to assess the roles of specific amino acid residues in virulence. This can be done by knocking out the native *Rv2633c* gene in *Mtb* and replacing it with genes that have been altered by site-directed mutagenesis to express variant proteins in macrophage and animal models of infection. These data can be correlated with solution studies of analogous variants of the *Mka* HLP *in vitro*. Thus, the structure of the *Mka* HLP and model of the *Mtb* HLP that are presented here will be relevant to understanding the role of the mycobacterial HLPs in mycobacterial infections in general.

Data Deposition

The co-ordinates of the structures presented here have been deposited in the Protein Data Bank: apo *Mka* HLP (6U3L), iron-bound *Mka* HLP (6Q09).

Abbreviations

HLP, hemerythrin-like protein; Mka, *Mycobacterium kansasii*; Mtb, *Mycobacterium tuberculosis*.

Author Contributions

All authors participated in the design of experiments, analysis of data and writing the paper. Z.M., J.A. and G.W. B. performed the experiments.

Funding

This research was funded in part by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under Federal Contract number HHSN272201700059C and by the National Institute of General Medical Sciences of the National Institutes of Health, under award R35GM130173 (V.L.D.).

Acknowledgements

The authors thank Yu Tang and Sandra Geden for technical assistance, Kyle Strickland and William T. Self for productive discussions, and the entire Seattle Structural Genomics Center for Infectious Diseases (SSGCID) team, including David M. Dranow for peer review of the structures and Thomas E. Edwards for critical review of the manuscript. The SSGCID internal ID for *Mka* HLP is MykaA.20209.a (www.ssgcid.org). Part of the research was conducted at the W.R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research

Biochemical Journal (2020) **477** 567–581
https://doi.org/10.1042/BCJ20190827

(BER) program located at Pacific Northwest National Laboratory (PNNL). Battelle operates PNNL for the U.S. Department of Energy.

Open Access

Open access for this article was enabled by the participation of University of Central Florida in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

References

1 Homolka, S., Niemann, S., Russell, D.G. and Rohde, K.H. (2010) Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathog.* **6**, e1000988 https://doi.org/10.1371/journal.ppat.1000988

2 Rohde, K.H., Abramovitch, R.B. and Russell, D.G. (2007) *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe.* **2**, 352–364 https://doi.org/10.1016/j.chom.2007.09.006

3 Abramovitch, R.B., Rohde, K.H., Hsu, F.-F. and Russell, D.G. (2011) aprABC: a *Mycobacterium tuberculosis* complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. *Mol. Microbiol.* **80**, 678–694 https://doi.org/10.1111/j.1365-2958.2011.07601.x

4 Cumming, B.M., Rahman, M.A., Lamprecht, D.A., Rohde, K.H., Saini, V., Adamson, J.H. et al. (2017) *Mycobacterium tuberculosis* arrests host cycle at the G1/S transition to establish long term infection. *PLoS Pathog.* **13**, e1006389 https://doi.org/10.1371/journal.ppat.1006389

5 Walters, S.B., Dubnau, E., Kolesnikova, I., Laval, F., Daffe, M. and Smith, I. (2006) The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Mol. Microbiol.* **60**, 312–330 https://doi.org/10.1111/j.1365-2958.2006.05102.x

6 Sassetti, C.M. and Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. *Proc. Natl Acad. Sci. U.S.A.* **100**, 12989–12994 https://doi.org/10.1073/pnas.2134250100

7 Ma, Z., Strickland, K.T., Cherne, M.D., Sehanobish, E., Rohde, K.H., Self, W.T. et al. (2018) The Rv2633c protein of *Mycobacterium tuberculosis* is a non-heme di-iron catalase with a possible role in defenses against oxidative stress. *J. Biol. Chem.* **293**, 1590–1595 https://doi.org/10.1074/jbc.RA117.000421

8 Stenkamp, R.E. (1994) Dioxygen and hemerythrin. *Chem. Rev.* **94**, 715–726 https://doi.org/10.1021/cr00027a008

9 Alvarez-Carreño, C., Alva, V., Becerra, A. and Lazcano, A. (2018) Structure, function and evolution of the hemerythrin-like domain superfamily. *Protein Sci.* **27**, 848–860 https://doi.org/10.1002/pro.3374

10 Chen, K.H.C., Chuankhayan, P., Wu, H.-H., Chen, C.-J., Fukuda, M., Yu, S.S.F. et al. (2015) The bacteriohemerythrin from *Methylococcus capsulatus* (Bath): crystal structures reveal that Leu114 regulates a water tunnel. *J. Biol. Inorg. Chem.* **150**, 81–89 https://doi.org/10.1016/j.jinorgbio.2015.04.001

11 Martins, L.J., Hill, C.P. and Ellis, Jr. W.R. (1997) Structures of wild-type chloromet and L103N hydroxomet *Themiste zostericola* myohemerythrins at 1.8 A resolution. *Biochemistry* **36**, 7044–7049 https://doi.org/10.1021/bi9630422

12 Alvarez-Carreno, C., Becerra, A. and Lazcano, A. (2016) Molecular evolution of the oxygen-binding hemerythrin domain. *PLoS One* **11**, e0157904 https://doi.org/10.1371/journal.pone.0157904

13 Isaza, C.E., Silaghi-Dumitrescu, R., Iyer, R.B., Kurtz, Jr. D.M. and Chan, M.K. (2006) Structural basis for O_2 sensing by the hemerythrin-like domain of a bacterial chemotaxis protein: substrate tunnel and fluxional N terminus. *Biochemistry* **45**, 9023–9031 https://doi.org/10.1021/bi0607812

14 Ruiz, J.C. and Bruick, R.K. (2014) F-box and leucine-rich repeat protein 5 (FBXL5): sensing intracellular iron and oxygen. *J. Inorg. Biochem.* **133**, 73–77 https://doi.org/10.1016/j.jinorgbio.2014.01.015

15 Lo, F.C., Hsieh, C.C., Maestre-Reyna, M., Chen, C.Y., Ko, T.P., Horng, Y.C. et al. (2016) Crystal structure analysis of the repair of iron centers protein YtfE and its interaction with NO. *Chemistry* **22**, 9768–9776 https://doi.org/10.1002/chem.201600990

16 Baugh, L., Phan, I., Begley, D.W., Clifton, M.C., Armour, B., Dranow, D.M. et al. (2015) Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis (Edinb)* **95**, 142–148 https://doi.org/10.1016/j.tube.2014.12.003

17 Li, C., Wen, A., Shen, B., Lu, J., Huang, Y. and Chang, Y. (2011) Fastcloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 https://doi.org/10.1186/1472-6750-11-92

18 Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 https://doi.org/10.1016/j.pep.2005.01.016

19 Buchko, G.W., Hewitt, S.N., Napuli, A.J., Van Voorhis, W.C. and Myler, P.J. (2011) Solution structure of an arsenate reductase-related protein, YffB, from *Brucella melitensis*, the etiological agent responsible for brucellosis. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67**, 1129–1136 https://doi.org/10.1107/S1744309111006336

20 Carter, P. (1971) Spectrophotometric determination of serum iron at the submicrogram level with a new reagent (ferrozine). *Anal. Biochem.* **40**, 450–458 https://doi.org/10.1016/0003-2697(71)90405-2

21 Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 https://doi.org/10.1107/S0907444909047337

22 Zwart, P.H., Afonine, P.V., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J. et al. (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 https://doi.org/10.1007/978-1-60327-058-8_28

23 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) *Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 https://doi.org/10.1107/S0021889807021206

24 Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478 https://doi.org/10.1107/S090744490903947X

25 Langer, G., Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 https://doi.org/10.1038/nprot.2008.91

26 Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 https://doi.org/10.1107/S0907444910007493
27 Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N. et al. (2018) Molprobity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 https://doi.org/10.1002/pro.3330
28 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R. et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucl. Acids Res.* **46**, W296–W303 https://doi.org/10.1093/nar/gky427
29 Benkert, P., Biasini, M. and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 https://doi.org/10.1093/bioinformatics/btq662
30 Wang, J., McIntosh, F., Radomski, N., Dewar, K., Simeone, R., Enninga, J. et al. (2015) Insights on the emergence of *Mycobacterium tuberculosis* from the analysis of *Mycobacterium kansasii*. *Genome Biol. Evol.* **7**, 856–870 https://doi.org/10.1093/gbe/evv035
31 Bloch, K.C., Zwerling, L., Pletcher, M.J., Hahn, J.A., Gerberding, J.L., Ostroff, S.M. et al. (1998) Incidence and clinical implications of isolation of *Mycobacterium kansasii*: results of a 5-year, population-based study. *Ann. Int. Med.* **129**, 698–704 https://doi.org/10.7326/0003-4819-129-9-199811010-00004
32 Smith, M.B., Molina, C.P., Schnadig, V.J., Boyars, M.C. and Aronson, J.F. (2003) Pathologic features of *Mycobacterium kansasii* infection in patients with acquired immunodeficiency syndrome. *Arch. Pathol. Lab. Med.* **127**, 554–560 https://doi.org/10.1043/0003-9985(2003)127<0554:PFOMKI>2.0.CO;2
33 Stenkamp, R.E., Sieker, L.C., Jensen, L.H., McCallum, J.D. and Sanders-Loehr, J. (1985) Active site structures of deoxyhemerythrin and oxyhemerythrin. *Proc. Natl Acad. Sci. U.S.A.* **82**, 713–716 https://doi.org/10.1073/pnas.82.3.713