doi:10.1016/j.jmb.2008.07.001 J. Mol. Biol. (2008) 383, 822–836

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Structure and Unusual Protein Chemistry of Hypoxic Response Protein 1, a Latency Antigen and Highly Expressed Member of the DosR Regulon in *Mycobacterium tuberculosis*

Miriam L. Sharpe[1], Chen Gao[1], Sharon L. Kendall[2], Edward N. Baker[1] and J. Shaun Lott[1]*

[1]*School of Biological Sciences and Maurice Wilkins Center for Molecular Biodiscovery, University of Auckland, Private Bag 92109, Auckland 1142, New Zealand*

[2]*Department of Pathology and Infectious Diseases, The Royal Veterinary College, Royal College Street, London NW1 0TU, UK*

Received 19 March 2008; received in revised form 30 June 2008; accepted 2 July 2008
Available online 9 July 2008

Edited by R. Huber

Mycobacterium tuberculosis adapts to cellular stresses such as decreased oxygen concentration, at least in part, by upregulation of the *do*rmancy *s*urvival regulon, which is thought to be important for the bacterium's ability to enter a persistent state in its human host. We have determined the structure of hypoxic response protein 1, a protein encoded by one of the most strongly upregulated genes in the dormancy survival regulon. Hypoxic response protein 1 is an example of a 'cystathionine-β-synthase-domain-only' protein; however, unlike other cystathionine-β-synthase domains, it does not appear to bind AMP. The protein is proteolytically sensitive at its C-terminus and contains two unexpected disulfide bonds, one of which appears resistant to reducing agents in solution and is, therefore, most likely buried in the protein and is not solvent-accessible. We show that the protein is secreted from the bacterium in hypoxic *in vitro* culture and does not accumulate in the bacterial cell wall. The biological function of the protein remains unclear, but we suggest that it may contribute to the modulation of the host immune response. The work reported advances our understanding of the chemistry and cell biology of this intriguing and potentially important protein, and establishes a structural framework for future functional and immunological studies.

Keywords: tuberculosis; dormancy survival; hypoxic response; latency; crystal structure

Introduction

The adaptation of the preferentially aerobic pathogen *Mycobacterium tuberculosis* to a decreased oxygen concentration is thought to be important for the bacterium's ability to enter an antibiotic-resistant nonreplicating persistent state in its human host.[1] Individuals infected by persistent bacteria are often asymptomatic for long periods of time, sometimes decades, before the bacteria re-emerge to cause disease. The control of gene expression in response to decreased oxygen tension has been shown to be governed, at least in part, by the dormancy survival transcriptional *r*egulator DosR.[2] DosR is also activated in response to exposure of the bacteria to nitric oxide, ethanol, and peroxide, but not to heat shock or cold shock.[3] DosR is itself the response regulator of a complex two-component system and integrates signals from two sensor histidine kinases, DosS and DosT, which are proposed to be differentially sensitive to redox state and hypoxia, respectively.[4]

One of the most strongly upregulated open reading frames in the DosR regulon is Rv2626c.[5,6] This

*Corresponding author. E-mail address: s.lott@auckland.ac.nz.

Present address: M. L. Sharpe, National Institute of Medical Research, The Ridgeway, Mill Hill East, London, UK.

Abbreviations used: HRP1, hypoxic response protein 1; CBS, cystathionine-β-synthase; AMPK, AMP-activated protein kinase; TFA, trifluoroacetic acid; PDB, Protein Data Bank; ADH, alcohol dehydrogenase; TNP, 2′,3′-*O*-(2,4,6-trinitrophenyl); rTEV, Recombinant Tobacco Etch Virus; PBS, phosphate-buffered saline; BSA, bovine serum albumin.

encodes a conserved hypothetical protein of unknown function, which we have previously termed hypoxic response protein 1 (HRP1).[7] HRP1 expression has been consistently upregulated (12- to 107-fold above resting levels) in several *in vitro* experimental systems upon the exposure of *M. tuberculosis* to ≤1% of normal oxygen saturation.[8–11] Increased expression of HRP1 was also visible under *in vitro* hypoxic conditions in a proteomics study using two-dimensional electrophoresis and metabolic labeling.[12] Transcription of HRP1 was significantly upregulated *ex vivo* in *M. tuberculosis* from within infected and activated murine macrophages, and *in vivo* in *M. tuberculosis* encapsulated in hollow fibers implanted in mice, in two different microarray studies.[13,14] HRP1 was also highly expressed by *M. tuberculosis in vivo* 15 or 21 days after mice were infected, as measured by quantitative real-time reverse transcriptase–PCR assays.[11,15] In addition to the effects of hypoxia, HRP1 transcription was upregulated *in vitro* in response to low nontoxic concentrations of NO,[11] which is consistent with the sensitivity of the DosS/DosR two-component regulatory system to both decreased O_2 and increased NO concentrations.[3] In contrast, no significant change in HRP1 expression was seen in oxygenated *M. tuberculosis* cultures grown under nutrient-depleted conditions over 96-h[16] or 75-day[17] time frames.

Sequence similarity analysis predicts HRP1 to be entirely composed of two cystathionine-β-synthase (CBS) domains, and hence to be a member of a widely distributed family of 'CBS-domain-only' proteins. CBS domains are found in all kingdoms of life (always as two or four tandem copies), either constituting the entire protein, as in HRP1, or attached to other types of globular protein domain, as in CBS, AMP-activated protein kinase (AMPK), and inosine monophosphate dehydrogenase.[18–21] CBS domains are also found in the intracellular portion of some integral membrane proteins, such as voltage-gated chloride channels[22,23] and, incidentally, in an *M. tuberculosis* protein of unknown function found in the same operon as HRP1 (Rv2625c).

CBS domains in eukaryotic multidomain proteins appear, in some cases, to have either an inhibitory or an activating role in those proteins, in response to the binding of AMP or other adenosine-containing ligands. A number of mutations in CBS-domain pairs found in multidomain proteins (e.g., CBS, AMPKγ2, and inosine monophosphate dehydrogenase) are known to cause a range of pathologies in humans. Scott *et al.* produced isolated CBS-domain pairs from these proteins recombinantly and showed that introduction of the pathology-associated mutations decreased their ability to bind adenosine-containing ligands.[24] In an earlier study by Janošik *et al.*, it was found that mutations in two residues from the CBS domains of CBS caused the enzyme to become insensitive to activation by *S*-adenosyl-L-methionine *in vitro*, suggesting that the CBS domains are required for functional *S*-adenosyl-L-methionine binding.[19] Adams *et al.* modeled the CBS domains of AMPKγ1 *in silico*.[25] Four residues comprising the predicted AMP-binding pocket from this model (Arg70, Arg152, Arg171, and His151) were mutated separately *in vitro*, and each of the mutations resulted in the loss of the AMP-dependent allosteric control of the enzyme. Recently, the ligand-bound structures of the *Schizosaccharomyces pombe* and mammalian AMPKγ and voltage-gated chloride channel CBS domains have been solved, revealing different modes of binding AMP and ATP.[26–29] It is unclear whether 'CBS-domain-only' proteins such as HRP1 have a function related to those of the larger multidomain proteins and are also able to bind adenosine-containing ligands.

In order to understand the role of the 'CBS-domain-only' protein HRP1 in the hypoxic response of *M. tuberculosis*, we have expressed the protein recombinantly, determined its three-dimensional structure using X-ray crystallography, characterized its solution properties, and carried out functional assays to investigate putative functions. Additionally, we have localized the protein within the bacterium using immuno-electron microscopy. Together, these studies have shown that HRP1 is a secreted protein that forms a tight dimer via a buried disulfide bond, and that it probably does not have the nucleotide-binding capability seen in some CBS domains from multidomain proteins. The physiological role of HRP1 in the bacterial hypoxic response remains unclear, but it is likely to be an important component of the bacterium's response to cellular stress and may be a useful antigenic marker of persistent bacteria in dormant tuberculosis.

Results

Solution properties of HRP1

When analyzed by SDS-PAGE, pure recombinant full-length His-tagged HRP1 (171 amino acid residues) migrated as two major bands, equivalent to the expected sizes of monomer and dimer species (approximately 19 and 38 kDa, respectively; Fig. 1). The presence on an SDS-PAGE gel of a band equivalent to the dimer indicated that the protein may form an unusually strong dimer in solution. SDS-resistant multimers are often associated with strongly hydrophobic membrane proteins such as bacterial porins,[31] rather than with soluble proteins such as HRP1. To investigate the stability of the HRP1 dimer, the protein was incubated with various reducing, denaturing, or acidic compounds, or was heated, and the ratio of the monomer-to-dimer bands was assessed using SDS-PAGE. The SDS-resistant dimer band persisted despite incubation with urea, several reducing agents, or trifluoroacetic acid (TFA) (Fig. 1a). As judged by this semiquantitative assay, the dimerization interaction only began to be disrupted after heating at 100 °C for 2 min or more, as demonstrated by the decreasing ratio of the intensity of the dimer band to the intensity of the monomer band in Fig. 1b. Size-exclusion chromato-

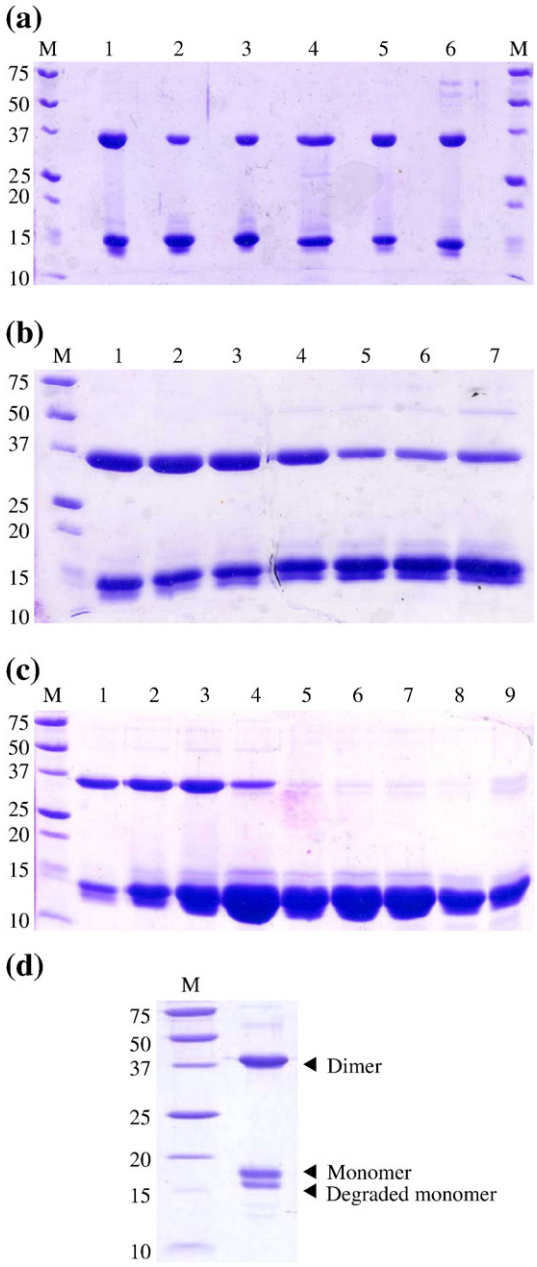


Fig. 1. SDS-PAGE analysis of HRP1 dimerization. (a) Recombinant HRP1 was incubated for 5 min at room temperature with the following additives: lane 1: alone; lane 2: in 7.8 M urea (a chaotropic denaturant); lane 3: in 150 mM DTT (a reducing agent); lane 4: in 150 mM tris (2-carboxyethyl) phosphine hydrochloride (a reducing agent stronger than DTT); lane 5: 10 mM mycothiol (a reducing agent found naturally in mycobacteria;[30] a gift from R. Fahey and G. Newton, University of California, San Diego, CA); lane 6: 0.042% vol/vol TFA. The protein was incubated in TFA as control for the incubation in mycothiol, which contained an equivalent concentration of TFA remaining from the preparation of the compound. Lanes marked M contain molecular mass markers (in kDa). (b) Recombinant HRP1 was heated at 100 °C for 0 s, 15 s, 30 s, 1 min, 2 min 30 s, 5 min, and 10 min (lanes 1–7). (c) Recombinant HRP1 was heated at 100 °C in the presence of 150 mM DTT for 0 s, 15 s, 30 s, 1 min, 2 min 30 s, 5 min, 10 min, 30 min, and 60 min (lanes 1–9). (d) Recombinant His-tagged HRP1 after storage at 4 °C for several weeks in the absence of protease inhibitors.

graphy demonstrated that the protein predominantly forms dimers in solution (data not shown) with the presence of some higher-order multimers, but with no evidence of a monomeric species in solution.

However, after storage for several weeks at 4 °C in the absence of any protease inhibitors, SDS-PAGE analysis of HRP1 revealed that a fraction of the full-length protein had been degraded, forming another band on an SDS gel just below the original monomer band (Fig. 1d). Analysis of the degraded protein by mass spectrometry showed that the C-terminal 16 amino acids were lost (data not shown). To investigate the effect of this truncation on the protein, a construct (HRP1-Δ128–143) that lacked the C-terminal 16 residues was produced. Unlike the full-length protein, HRP1-Δ128–143 did not form SDS-resistant dimers on SDS-PAGE (Fig. 2, lane 2), although size-exclusion chromatography revealed that it remains dimeric in solution. This suggests that the 16-residue C-terminal tail of the full-length protein (IVQFV-KAICSPMALAS) is necessary for it to form the observed SDS-resistant dimers, but is not required for dimerization *per se*.

Protein crystallization and structure solution

Full-length HRP1 could not be readily or reproducibly crystallized.[7] As flexible or disordered regions of proteins can prevent crystallization, it was reasoned that the easily degraded 16-residue C-terminal tail region of HRP1 may be inhibiting the crystallization of the full-length protein. After cleavage of the N-terminal His tag, both native and selenomethionine-substituted forms of HRP1-Δ128–143 readily yielded diffraction-quality crystals in a number of conditions, which were used to solve

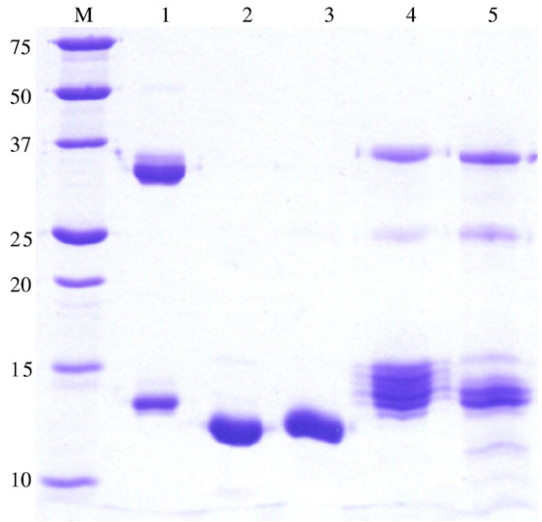


Fig. 2. The role of the C-terminal tail in the dimerization of HRP1. SDS-PAGE analysis of wild-type and mutant HRP1. Lane 1: wild-type HRP1; lane 2: HRP1-Δ128–143; lane 3: HRP1-C136A; lane 4: HRP1-C14A; lane 5: HRP1-C39A. The lane marked M contains molecular mass markers (in kDa).

Table 1. Data collection statistics

	Native HRP1-Δ128–143	SeMet HRP1-Δ128–143
Crystal parameters		
Space group	$P2_12_12_1$	$C2$
Cell parameters		
a (Å)	40.56	77.45
b (Å)	50.94	41.06
c (Å)	128.19	78.40
β (°)		108.41
Data collection		
X-ray source	Home (Rigaku RU-H3R)	Advanced Light Source (Beamline 5.0.2)
Wavelength (Å)	1.5418	0.9793
Resolution (Å)	30.00–2.00 (2.07–2.00)	50.00–1.60 (1.66–1.60)
Reflections (measured/unique)	182,705/18,573	199,360/30,300
Multiplicity[a]	9.8	6.6
Completeness (%)	98.2 (96.8)	97.3 (84.6)
$\langle I\rangle/\langle\sigma I\rangle$	19.7 (4.1)	15.6 (3.0)
R_{merge}[b]	0.116 (0.547)	0.118 (0.329)
Mosaicity (°)	0.38	0.39

Values in parentheses are for the outermost shell of data.

[a] Multiplicity calculated with Friedel-related reflections separated for SeMet data.

[b] $R_{merge} = \frac{\sum |I - \langle I\rangle|}{\sum I}$.

the structure using the single-wavelength anomalous diffraction method.

The structure of HRP1-Δ128–143 was determined at 1.6 Å resolution using data collected from a SeMet-substituted crystal and refined to final R_{cryst} and R_{free} values of 22.2% and 23.4%, respectively. The SeMet structure was then used in molecular replacement to solve the structure of native HRP1-Δ128–143, which was refined to 2.0 Å resolution, with final R_{cryst} and R_{free} values of 20.7% and 23.8%, respectively. Data collection statistics are summarized in Table 1, and phasing and refinement statistics are summarized in Table 2.

Both of the final models contain two protein molecules in the asymmetric unit. The Matthews coefficients (V_m) for the native and SeMet structures are 2.3 and 2.0 Å^3/Da, respectively, and the estimated solvent contents are 46.8% and 39.3%. The SeMet model consists of residues 2–124 for chain A, and residues 2–123 for chain B, together with 136 water molecules. The final model for native HRP1-Δ128–143 consists of residues 2–74 and 78–124 for chain A, and residues 2–123 for chain B, together with 204 water molecules and two Ni^{2+}. The number of water molecules, especially in the SeMet structure, is unusually low for a model with 266 residues in the asymmetric unit,[32] but no significant ($\geq 3\sigma$) peaks remained in difference maps in which water molecules could be placed. In Ramachandran plots produced by PROCHECK,[33] 94.4% and 93.5% of residues from the native and SeMet structures, respectively, are in the most favored regions, and 5.6% and 6.5% are in additionally allowed regions. No residues in either structure are in generously allowed regions or disallowed regions (Table 2).

Molecular structure of HRP1-Δ128–143

The crystal structure of HRP1-Δ128–143 reveals a monomer composed of two CBS domains (Fig. 3), as predicted from the sequence. CBS domains were originally defined by sequence conservation.[18] However, one copy of the conserved sequence does not constitute a domain by itself, in the sense of a discrete, autonomously folded, globular structural unit, and all known structures are of CBS-domain pairs, which have also been referred to as Bateman domains.[18]

Each CBS domain in the HRP1-Δ128–143 structure adopts the same basic structure as that described in Bateman's original definition: a β-sheet of three or four β-strands, with helices (in this case three) packed against one face of the sheet. The CBS domain that contains both the N-terminus and the C-terminus of the protein is referred to here as CBS2, and the other CBS domain is referred to as CBS1.

Table 2. Phasing and refinement statistics

Phasing SeMet HRP1-Δ128–143		
Molecules in asymmetric unit	2	
Selenium sites used in phasing	10 of 12 possible sites present	
Measures of solution quality (SHELXD) (%)		
Correlation coefficient	50.94	
Weak correlation coefficient	33.32	
Patterson Figure of Merit	14.45	
	Native HRP1-Δ128–143	SeMet HRP1-Δ128–143
Refinement		
Resolution range (Å)	20.5–2.0	38.0–1.6
Number of reflections used (% of theoretical)	21,428 (95.2)	66,279 (90.2)
Number of reflections in test set (% of working)	1071 (4.8)	6466 (9.8)
R_{cryst}/R_{free}[a]	0.206/0.235	0.222/0.233
Values for outermost shell of data	(0.313/0.365)	(0.329/0.328)
Number of protein atoms	1813	1760
Number of water molecules	204	136
Number of nonprotein and nonwater atoms	2 (Ni^{2+})	0
B-factor		
Average (Å^2)	22.4	26.8
RMSD from ideality		
Bond length (Å)	0.005	0.004
Bond angle (°)	1.20	1.20
Residues in the Ramachandran plot [n (%)]		
Most favored regions	201 (94.4)	203 (93.5)
Additionally allowed regions	12 (5.6)	14 (6.5)
Generously allowed regions	0 (0)	0 (0)
Disallowed regions	0 (0)	0 (0)

The cross-validated R-factor, R_{free}, is calculated in the same way, but using only a randomly selected subset of data that have been excluded from the refinement calculations.

[a] $R_{cryst} = \frac{\sum |F_{obs}| - |F_{calc}|}{\sum |F_{obs}|}$.

The overall secondary structure of the monomer consists of seven β-strands in two mixed β-sheets and six α-helices, three packed against the outer face of each β-sheet (Fig. 3a). Four β-strands (β1, β5, β6, and β7) are found in the CBS2 β-sheet, and three strands (β2, β3, and β4) are found in the CBS1 β-sheet. The faces of the two β-sheets that are not associated with the α-helices face each other, forming a β-sandwich. This face-to-face β-sheet interaction between the CBS domains is facilitated by hydrophobic side chains from each sheet, including Pro103, Ile105, Leu110, and Ile113 from CBS2, and Pro37, Leu45, and Met48 from CBS1. The hydrophobic nature of this interface highlights the fact that two sequence-defined CBS domains are required to form a structurally stable domain. On either side of the monomer, running almost perpendicular to the faces of the two β-sheets and securing the face-to-face sheet interaction are two 'straps' that have little regular structure. The N-terminal 'strap' consists of the residues between β1 and β2 (residues 11–14), and the C-terminal 'strap' consists of the residues between α4 and β5 (residues 75–80). The C-terminal strap is poorly defined in one monomer of the native structure, with no interpretable electron density for residues 75–77, suggesting that this is a flexible region.

As implied by sequence conservation, HRP1 contains a repeat in structure that probably originated in a gene duplication event (Fig. 3b). A structural overlay of residues 1–65 and 66–123 (RMSD of 2.06 Å over 285 equivalent atoms) shows that the two CBS domains share a similar conformation (Fig. 3c). Most secondary structure elements are conserved across the two regions: helices α2 and α5 are equivalent, as are strands β3 and β6, strands β4 and β7, helices α3 and α6, and helices α1 and α4. However, there is no equivalent strand for β1, which in the complete structure would have run antiparallel with β4.

Unexpectedly, a solvent-accessible intramolecular disulfide bond was found in each monomer between Cys14 and Cys39, which connects the N-terminal strap with the β3 strand of CBS1. The disulfide bond is not conserved in other CBS-domain-containing structures in the Protein Data Bank (PDB). The presence of a disulfide bond is unusual for an intracellular protein, and the cytosol of mycobacteria is presumably a reducing environment,[30] suggesting that the protein may be secreted *in vivo*.

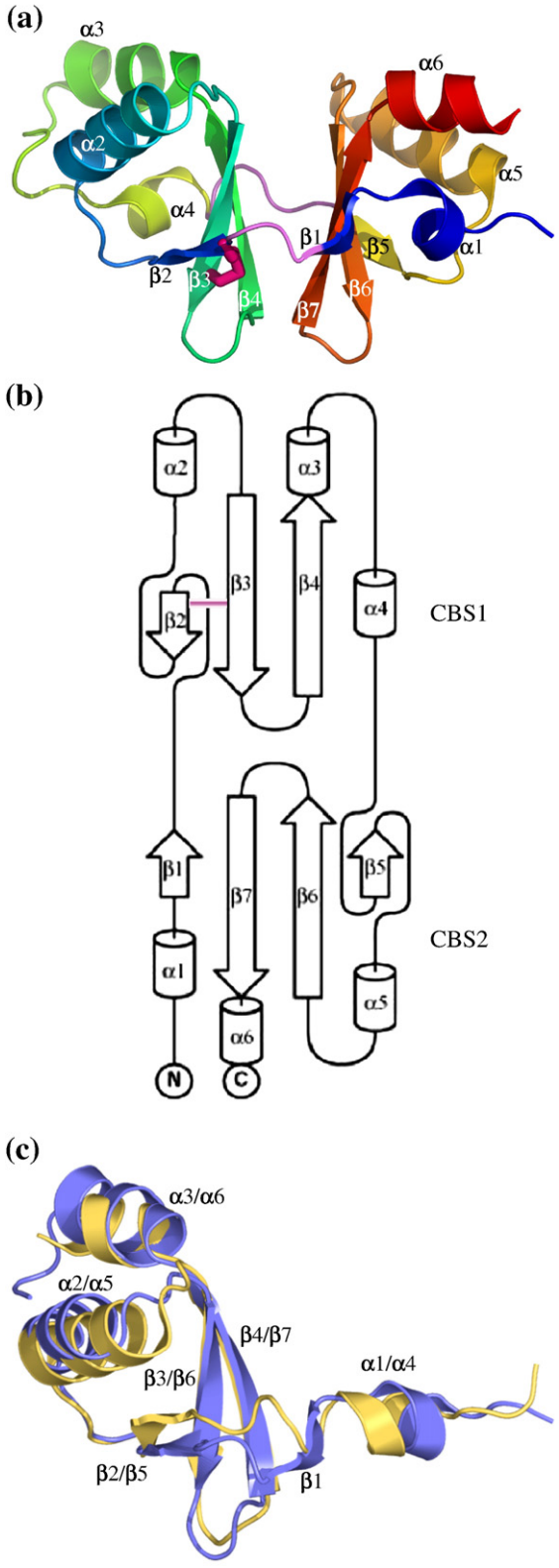


Fig. 3. The structure of HRP1-Δ128–143. (a) A ribbon diagram of the monomer is shown in blue to red from the N-terminus to the C-terminus. Secondary structure elements are labeled, and the side-chain atoms of the intramolecular disulfide bond between C14 and C39 are shown in a thick magenta stick representation. The 'strap' regions connecting the two CBS domains have been highlighted in pink. This and subsequent figures were made using PyMOL.[34] (b) A topology diagram of the secondary structure of the HRP1-Δ128–143 monomer illustrating the duplication of the CBS domains. The position of the intramolecular disulfide bond is marked as a line in magenta. The diagram was created using TopDraw.[35] (c) A ribbon representation of residues 1–65 (blue) superposed onto residues 66–123 (yellow) from the crystal structure of HRP1-Δ128–143 illustrating the structural conservation between the two CBS domains.

The roles of Cys14, Cys39, and Cys136 in stability and dimerization

In order to investigate the role of the observed disulfide bond in the stability of the protein, two site-specific mutants were made, changing each of Cys14 and Cys39 to an alanine residue (HRP1-C14A and HRP1-C39A). A third construct— which contained a single mutation within the C-terminal tail, changing the third cysteine residue in the monomer Cys136 into an alanine residue (HRP1-C136A)—was also created. The mutant proteins were expressed and purified in the same manner as the native protein and were assessed for their ability to form SDS-resistant dimers.

All of the proteins were still dimeric in solution, as judged by analytical size-exclusion chromatography (data not shown). Both the C14A and the C39A mutants were able to form SDS-resistant dimers like the native protein, but were less stable in solution, undergoing more rapid degradation during purification and when stored at 4 °C (Fig. 2, lanes 4 and 5). This shows the importance of the disulfide bond for stability. In contrast, the C136A mutant did not form SDS-resistant dimers (Fig. 2, lane 3), indicating that Cys136 is responsible for the C-terminally mediated strong intermolecular association, most likely by forming an intermolecular disulfide bond. The resistance of the full-length native protein dimers to reducing agents indicates that this disulfide bond is probably protected from the solvent by the tertiary or quaternary structure of the protein. To test whether the SDS-resistant dimers were indeed mediated by an intermolecular disulfide bond, samples of native protein were incubated at 100 °C in the presence or in the absence of 150 mM 1,4-dithiothreitol (DTT). In the absence of DTT, the dimer band diminished only after heating for 2.5 min or more and was still clearly present after 10 min at 100 °C (Fig. 1b). However, in the presence of DTT, the dimer band was diminished after heating at 100 °C for 1 min and had completely disappeared after 2.5 min (Fig. 1c), indicating that thermal denaturation of the protein allows the otherwise shielded disulfide bond to be reduced under these circumstances.

In order to confirm the presence of a disulfide bond mediated by Cys136, we analyzed recombinant full-length HRP1 using a mass spectrometry approach. Full-length recombinant HRP1 was subjected to complete digestion with trypsin, and the tryptic peptides were then identified using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (Table 3). Peptides corresponding to the tryptic fragments encompassing both Cys14 (Asp6-Arg29) and Cys39 (Glu30-Arg44) were identified. An additional fragment corresponding to these two peptides joined via a disulfide bond was also observed, thus verifying both the existence of the Cys14-Cys39 disulfide bond observed in the crystal structure and the ability of this methodology to detect disulfide linkages. As mentioned previously, the full-length protein was proteolytically sensitive at its C-terminus, and peptides representing the intact C-terminal tryptic fragment were not observed. However, a fragment encompassing Cys136 (Ala134-Met139) was observed in the mass spectrum, as was a species corresponding to two copies of this peptide joined via a disulfide bond, hence experimentally confirming the existence of a Cys136-Cys136 intermolecular disulfide in full-length recombinant HRP1.

Table 3. Mass spectrometry analysis of HRP1

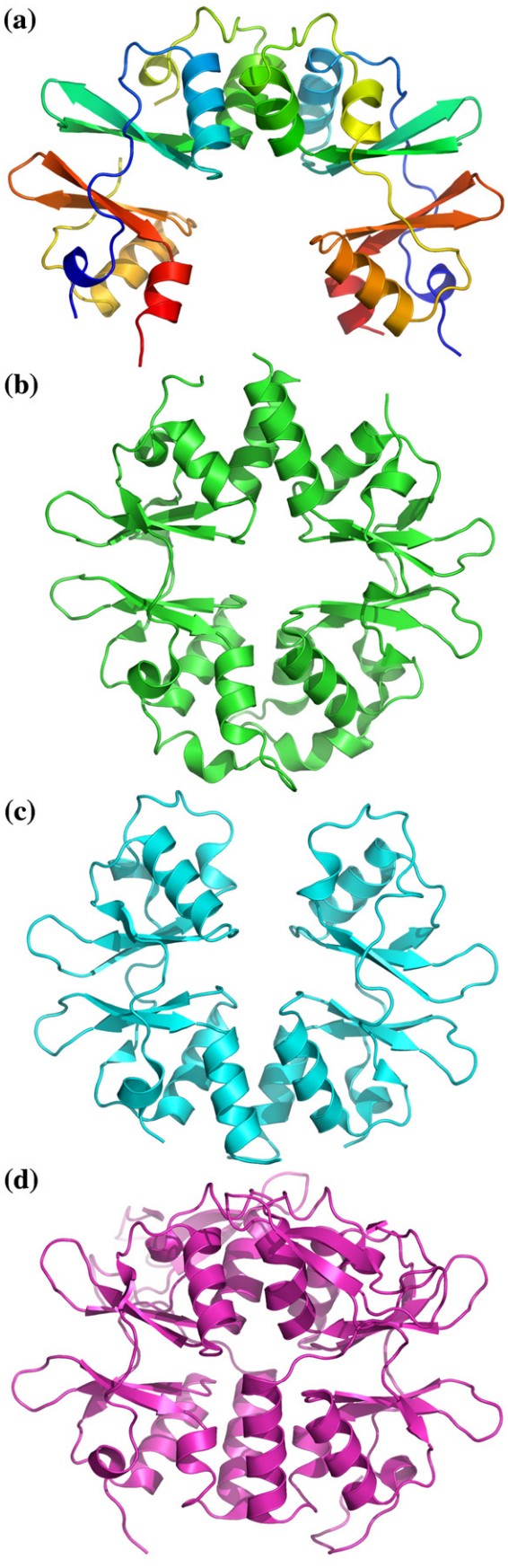
Tryptic peptide sequence	Expected mass (Da)	Observed mass (Da)
6-DIMNAGVTCVGEHETLTAAAQYMR-29	2581.9	2581.3
30-EHDIGALPICGDDDR-44	1625.7	1625.9
6-DIMNAGVTCVGEHETLTAAAQYMR-29 | 30-EHDIGALPICGDDDR-44	4205.6	4206.2
134-AICSPM-139	549.2	550.7
134-AICSPM-139 | 134-AICSPM-139	1096.4	1092.7

Dimerization interface

Three different potential dimer interfaces between the A and B monomers are seen in both the native HRP1-Δ128–143 crystal and the SeMet crystal. Of these three, the one that is most likely to be that present in solution is that with the most extensive buried surface area and the highest shape complementarity (Fig. 4a), as calculated using the program SC.[37] This choice of dimer interface is supported by its similarity to dimers seen in the structures of other CBS-domain-only proteins from *Bacillus subtilis*, *Thermotoga maritima*, and *Thermoplasma acidophilum* (PDB accession codes 1YAV, 1VR9 and 1PVM; Fig. 4b–d). A common dimerization face is used in all four cases, although there is evidently some variation in the contacts made between the monomers in the different proteins. In the HRP1-Δ128–143 dimer interface, only helices from the CBS1 domains of each monomer make contact with each other (Fig. 4a). In contrast, in the other protein dimers, the interfaces involve either the helices from the equivalents of the CBS2 domains, as in 1YAV (Fig. 4c); the helices in both CBS domains, as in 1VR9 (Fig. 4b); or helices from CBS1 together with additional C-terminal helices, which are not part of the conserved CBS-domain fold, as in 1PVM (Fig. 4d).[36]

Comparison with structurally related proteins

A search of the currently available protein structures in the PDB using the program SSM[38] found that the closest structural homologues of HRP1-Δ128–143 are all proteins of unknown or partially characterized function, mainly derived from structural genomics projects, as described above. These share only 19–24% sequence identity with HRP1-Δ128–143, although they display essentially conserved structures: superpositions of HRP1-



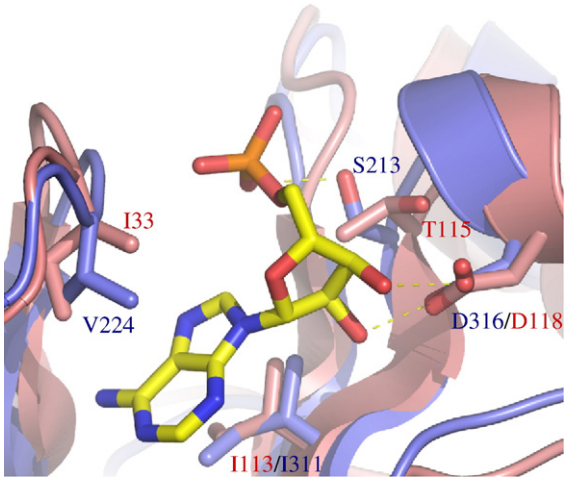


Fig. 5. Rv2626c and AMPKγ CBS domains show conservation around the latter's AMP binding site. The important ligand-interacting side chains in the AMPKγ nonexchangeable AMP binding site, as identified by Xiao *et al.*,[26] are shown along with their structural equivalents in HRP1. AMPKγ (PDB ID 2V8Q) is shown and labeled in blue; HRP1 is shown in pink and labeled in red.

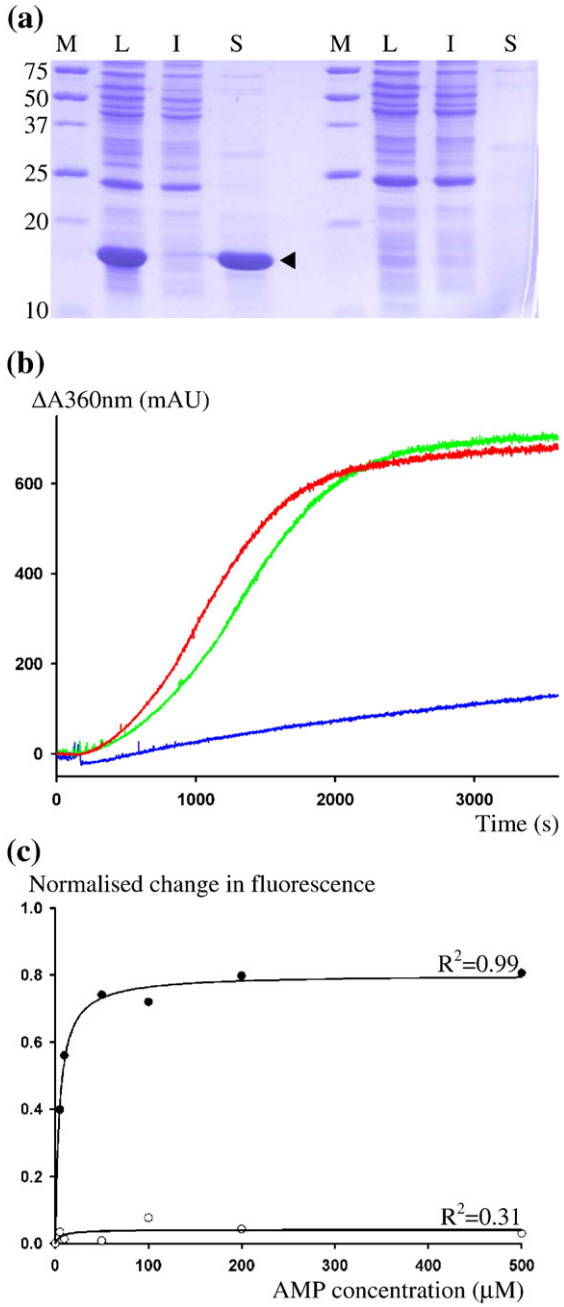
Δ128–143 onto the other proteins show that 109–115 residues can be matched, with an RMSD of 1.71–2.31 Å in C^{α} positions. All of these homologous structures contain the major features seen in the original definition of the CBS domain.[18] However, the structures of two proteins of well-characterized function that contain CBS domains have also recently been solved: the regulatory γ-subunit of the eukaryotic AMP kinase (AMPKγ)[26,27,29] and the cytosolic domain of the bacterial magnesium transporter MgtE.[39]

The paired CBS domains from AMPKγ share an essentially conserved architecture and dimeric arrangement with those from HRP1 and can be superposed with RMSD values of 1.67 and 2.42 Å, respectively, for monomer A (107 residues) and monomer B (113 residues) of HRP1. However, there is once again variability across the dimer interface, with the CBS-domain pairs being arranged in a more open configuration in HRP1 than in AMPKγ. Mammalian AMPKγ has three AMP binding sites, one of which does not exchange with ATP,[26] whereas AMPKγ from yeast contains only a single AMP binding site, which is structurally equivalent

Fig. 4. Variation in the dimerization interfaces of CBS-only proteins. Ribbon diagrams of (a) HRP1, colored as in Fig. 3a, illustrating that contacts between the CBS1 domains dominate the dimerization interface. (b) CBS-only protein YkuL from *B. subtilis* (PDB ID 1YAV) in green, which shows contributions from both CBS domains to the dimer interface. (c) CBS-only protein TM0892 from *Thermot. maritima* (PDB ID 1VR9) in cyan, in which the dimerization is mediated by the CBS2 domains. (d) CBS-only protein TA0289 from *Thermop. acidophilum*[36] (PDB ID 1PVM) in magenta, where the dimer interface is composed of the CBS1 domains and additional C-terminal helices.

to the nonexchanging site in the mammalian protein.[27] Superposition of mammalian AMPKγ with HRP1 shows that, whereas the exchangeable AMP binding sites from AMPKγ are poorly conserved in HRP1, key AMP-interacting residues from the nonexchangeable site are conserved in HRP1 (Fig. 5). Hydrophobic residues that pack on either side of the adenine ring in AMPKγ are seen in very similar conformations in HRP1, as is the aspartate residue (Asp316) that in AMPKγ forms bidentate hydrogen bonds with the 2′ and 3′ hydroxyl groups of the ribose ring. In other words, this region appears to be a potentially functional AMP binding site in HRP1.

In the magnesium transporter MgtE, the CBS domains form part of a Mg^{2+}-triggered conformational switch, with Mg^{2+} binding between the CBS domains and additional cytosolic domains of the protein in order to open and close the transmembrane Mg^{2+} channel.[39] Although one of the ligands used in Mg^{2+} binding in MgtE (Asp247) is structurally conserved in HRP1 (Asp118), there is no opportunity to form an equivalent Mg^{2+} binding site in HRP1 without the involvement of other proteins. However, the dimerization face seen in HRP1 and AMPKγ is also conserved in MgtE, and the orientation of the monomers with respect to each other is altered by ~20° during the Mg^{2+}-mediated conformational change, again illustrating the flexibility that exists across this conserved dimerization interface.



Functional prediction and analyses

Analysis of the HRP1-Δ128–143 structure was carried out using the ProFunc server[40] to attempt to predict a molecular function for the protein. None of the ProFunc analyses on HRP1-Δ128–143 delivered any more clues on a possible biochemical function for HRP1. However, since HRP1 is highly upregulated during exposure to cellular stress, we hypothesized that it may display chaperone activity like the *M. tuberculosis* α-crystallin homolog Acr,[8,41] which is also controlled by the DosR transcription factor. Possible chaperone activity was investigated using two assays, both testing the ability of HRP1 to prevent the thermal denaturation of other proteins. In the first assay, cell lysate taken from *Escherichia coli* cells expressing HRP1-Δ128–143 was heated to 76 °C for 10 min to determine whether its presence prevented the other proteins in the lysate from denaturing. Although HRP1 remains soluble after heat treatment, it fails to rescue any other proteins in the cellular lysate from denaturation (Fig. 6a). In the second,

Fig. 6. HRP1 does not display chaperone or AMP binding activity. (a) SDS-PAGE analysis of the thermal stability of HRP1 in an *E. coli* lysate (left-hand lanes) and a negative control of lysate without HRP1 (right-hand lanes). M, molecular mass markers; L, *E. coli* lysate sample before heating; I, insoluble fraction after heating; S, soluble fraction. HRP1 is marked with an arrowhead. (b) Time-dependent thermal denaturation curves of ADH alone (4 μM; green line) and in the presence of either HRP1 (20 μM; red line) or bovine eye lens α-crystallin (0.3 μM; blue line). Protein aggregation was measured by following the change in scattering at 360 nm. Molarity values were corrected for the molecular mass of the multimer formed in solution by each protein, as described by Yuan *et al.*[41] Therefore, a molecular mass of 360 kDa was assumed for α-crystallin (the multimerized form of the 16-kDa protein), a molecular mass of 80 kDa was assumed for ADH, and a molecular mass of 31.8 kDa was assumed for HRP1 (the dimerized form of the 15.9-kDa protein). (c) The normalized decrease in intrinsic protein fluorescence as a function of AMP concentration was measured for HRP1 (open circles) and PAE2307 (closed circles), a known AMP binding protein.[42] Ligand binding curves were fitted using SigmaPlot v.10, using a one-site saturation model $\Delta F = \frac{\Delta F_{max}[AMP]}{K_d + [AMP]}$. R^2 values from curve fitting, indicating the goodness of fit to the model in each case, are shown.

more quantitative assay, the ability of HRP1 to prevent the thermal denaturation of equine liver alcohol dehydrogenase (ADH) was measured using light scattering, a method used previously to demonstrate the ability of Acr to act as a chaperone.[41] Although HRP1 is not itself denatured by prolonged incubation at 48 °C (data not shown), it has no effect on the thermal denaturation of ADH at this temperature, unlike the chaperone α-crystallin (Fig. 6b).

As described above, some other CBS domains are known to bind adenosine-containing ligands. We investigated the possibility that HRP1 may also bind AMP using fluorescence spectroscopy, looking for changes to the intrinsic HRP1 fluorescence upon the addition of AMP. Although HRP1 does not contain any tryptophan residues, which usually dominate the fluorescence of proteins, it does contain three tyrosine residues and one phenylalanine residue from which fluorescent emission can also be observed.[43] In particular, the three tyrosine residues are located near the possible AMP binding pocket between the CBS domains in each monomer, and hence are likely to be sensitive to ligand binding. The intrinsic fluorescence of HRP1 as a function of the concentrations of added AMP is shown plotted in Fig. 6c, along with that of PAE2307—a putative nucleoside kinase from *Pyrobaculum aerophilum* that is known to have AMP binding activity[42]—as positive control. No significant changes in intrinsic fluorescence emission from HRP1 were observed with the addition of AMP, distinct from the saturable binding curve seen for PAE2307. Therefore, it appears that HRP1 does not bind AMP, at least not in any way that affects the fluorescent residues in the protein. As an alternative approach, the fluorescent AMP analogue 2′,3′-*O*-(2,4,6-trinitrophenyl) (TNP) AMP was also tested for its ability to bind to HRP1. In agreement with the intrinsic protein fluorescence data, no change in TNP-AMP fluorescence was observed in the presence of HRP1 (data not shown), indicating that TNP-AMP does not bind to the protein. While it is impossible to rule out that this failure to detect any AMP binding is due to the methodology chosen, the inability to bind to AMP (as measured by thermal shift circular dichroism spectroscopy), despite the apparent conservation of the putative binding cleft, is also observed for another CBS-domain protein, TA0289.[36]

Immunolocalization of HRP1

The presence of two disulfide bonds in HRP1 implies that the protein may be secreted *in vivo*. However, the sequence of the protein does not contain a signal peptide. We therefore set out to identify the cellular location of HRP1 by immunogold-labeling cultured *M. tuberculosis* cells with a specific polyclonal antibody raised against recombinant HRP1. Wild-type *M. tuberculosis* H37Rv cells, which had been grown under aerobic conditions and then left to settle and become hypoxic, showed a strong and specific pattern of labeling with a polyclonal anti-HRP1 antibody (Fig. 7a; 149 particles/μm^2), which was absent in cells treated with preimmune serum from the same rabbit (Fig. 7b; 26 particles/μm^2). Labeling was observed both in and immediately around the cells, in a pattern consistent with a secreted protein, but was not concentrated within the cells or clustered at the cell wall, as was observed previously for Acr,[44] indicating that HRP1 did not seem to be cell-wall-specific.

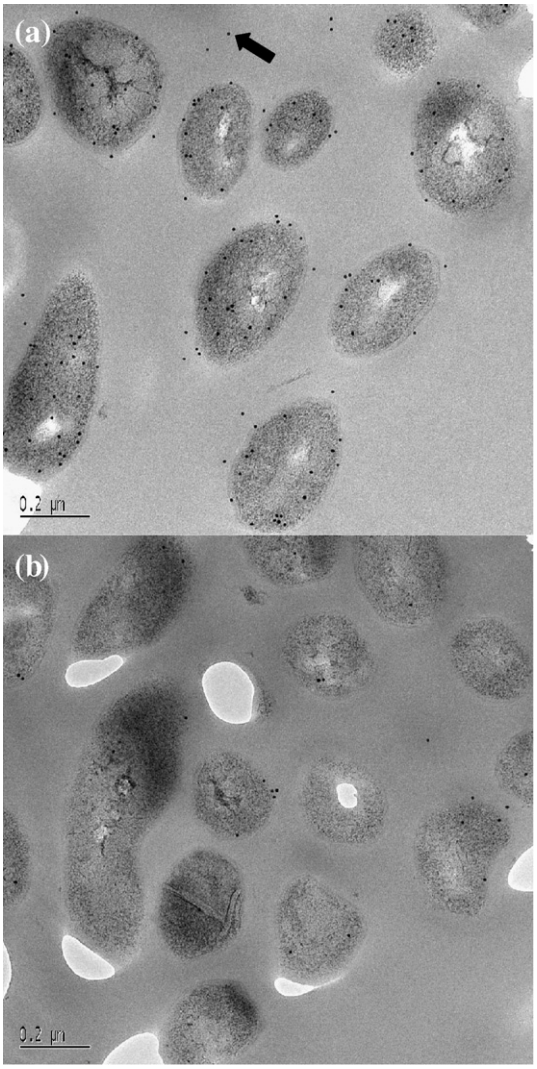


Fig. 7. Immunogold electron microscopy localization of HRP1. (a) HRP1 expression in wild-type *M. tuberculosis* H37Rv was visualized using rabbit polyclonal anti-HRP1 and a gold-labeled secondary antibody. The protein was found to be associated with the bacteria and also in the surrounding medium; an example of the latter is indicated with an arrow. (b) Lack of significant labeling using preimmune serum from the same rabbit.

Discussion

Structure

The determination of the structure of the HRP1-Δ128–143 monomer has shown that it is composed

of a tandem pair of CBS domains. The topology of the monomer is typical of that seen in other CBS-domain pair structures, with each CBS domain consisting of a β-sheet of three or four β-strands, with three helices packed against one face of the sheet, and with the other face of the sheet forming the interaction face with the second CBS domain. The determination of the crystal structure of HRP1-Δ128–143 and further biochemical investigation of the full-length protein have revealed unexpected chemical features of the protein, including an intramolecular disulfide bond and the ability to form very stable dimers, facilitated by a second intermolecular disulfide bond.

We have established that HRP1 is a dimer in solution, and that the most likely mode of dimerization is with the two helical bundles from each monomer contacting the helical bundles from the other monomer (CBS1-to-CBS1 and CBS2-to-CBS2). This conclusion is based on buried surface area and shape complementarity analyses on the dimer interfaces found in the HRP1-Δ128–143 crystals, and on comparison to dimer interfaces seen in the crystal structures of structural homologues.

A model for the structure of the HRP1 C-terminus

The 16 C-terminal residues removed from HRP1 to produce HRP1-Δ128–143 are predicted by a range of different secondary structure prediction algorithms to form a helical structure. The equivalent C-terminal region of the CBS-domain-containing protein TA0289 from *Thermop. acidophilum* also forms a helix, suggesting that HRP1 may contain an additional α-helix as part of the C-terminal tail, likely to be arranged in an orientation similar to that of the TA0289 C-terminal α-helix (labeled as α8 by Proudfoot *et al.*[36]), wrapped back against the CBS2 domain in the dimerization interface (Fig. 8a).

When the C-terminal sequences of HRP1 and TA0289 are aligned, Cys136 from HRP1 is in an equivalent position to Arg138 from TA0289. In the structure of the TA0289 dimer, the two Arg138 residues are positioned on the C-terminal end of each C-terminal helix, facing the opposite monomer in the dimer, with their C^{α} atoms 18 Å apart.

Assuming that the full-length HRP1 protein forms a dimer in the same manner as TA0289, the two Cys136 residues of HRP1 may be close enough to each other to form a disulfide bond between the helices. In TA0289, these two helices lie in a 'V' shape, so that their backbones are 6.6 Å apart at their N-terminal ends and 18 Å apart at their C-terminal ends, forming an approximately 30° angle between them. In order for the helices to be close enough for disulfide bonding to occur at the C-termini (a C^{α}–$C^{\alpha\prime}$ distance of 4.8–6.8 Å[46]), the helices would need to be repositioned to be approximately parallel, leaving the N-termini in the same relative positions. The final residue in HRP1, Ser143, is in an equivalent position to Leu145 from TA0289, which therefore approximates where the C-terminus of HRP1 may

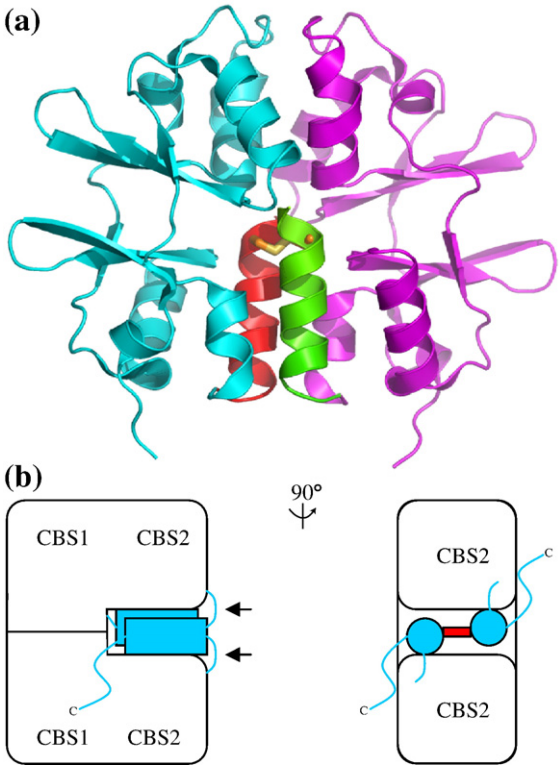


Fig. 8. Model for the C-terminal tail of HRP1. (a) Proposed model for the structure of the HRP1 C-terminal tail, produced by comparative modeling using Modeller.[45] The proposed dimer is shown in the same approximate orientation as in Fig. 4. The two monomers are in cyan and magenta, and the C-terminal helices (modeled on the structure of the TA0289 C-termini) are shown in red and green. The intermolecular disulfide bond between Cys136 in each monomer is shown as a thick yellow stick. (b) Schematic representation. The C-terminal helices are in blue, and the intermolecular disulfide bond is in red. The arrows indicate putative proteolytic cleavage sites.

be positioned. This repositioning of the C-terminal helices would require some flexibility in the CBS-domain dimerization interface. However, as noted above, variability in the region is already apparent when different members of the structural family are compared.

The consequence of this repositioning would be that the CBS2-to-CBS2 interface in the full-length HRP1 dimer would consist mainly of the two putative C-terminal helices packing next to each other, close enough for a disulfide bond to form between them via the Cys136 residues, which would be situated at the C-terminal end of the helices facing towards each other. This model would be consistent with the unusual protein chemistry observed for the intersubunit disulfide bond in HRP1, as the bond would then be protected from the solvent, mainly by the bulk of each monomer, and on either side of the dimer, by the helices from which they emerge (Fig. 8a).

However, the C-terminal tail of the full-length protein was observed to degrade during storage at 4 °C. If the model proposed above is correct, the loop

immediately N-terminal to the C-terminal helices (residues 128–131) is presumably susceptible to proteolysis. Cleavage at this site may result in the linked C-terminal helices dissociating from the rest of the dimer when exposed to SDS, allowing the dimer to dissociate as well.

Biological significance

The sequence of HRP1 does not contain any recognizable secretion signals, implying that it is secreted via a pathway independent of the well-characterized Sec and Tat secretion systems. This may mean that HRP1 is secreted from the cell via a dedicated mechanism analogous to the recently described 'type VII' secretion system involved in the export of the Esx antigens.[47] Initial immunological analyses of proteins whose expression is upregulated by DosR have indicated that they may act as antigens specific for latent infection in humans and mice, and that HRP1 in particular is strongly antigenic in mice.[48] Although the biochemical and biological functions of HRP1 remain unclear, both its potential utility as a marker of persistent infection and its importance as a target for protective vaccination are strong. It may be that HRP1 is involved in the modulation of the host immune response or macrophage behavior in a manner analogous to that proposed for the secreted EsxA/EsxB antigens.[49] The work reported represents an advancement in our understanding of the chemistry of this intriguing and potentially important protein, and establishes a structural framework for future functional and immunological studies.

Materials and Methods

DNA cloning

PCR amplifications of the full-length HRP1 open reading frame and the HRP1-Δ127–143 truncated construct were carried out with Platinum® *Pfx* polymerase (Invitrogen) using *M. tuberculosis* H37Rv genomic DNA as template. The primers used to produce the full-length construct are listed in Table 4. The full-length PCR product was digested with the restriction enzymes BamHI and XhoI and ligated into pProEX (Invitrogen) to produce pProEX-HRP1, and the truncated construct PCR product was inserted into pET151/D-TOPO® (Invitrogen) to produce pET151-HRP1-Δ128–143. These plasmids add a Recombinant Tobacco Etch Virus (rTEV) protease-cleavable 28-residue or 33-residue N-terminal His tag to the constructs, respectively.

Mutations were introduced into the full-length pProEX-HRP1 expression vector to produce the HRP1-C136A, HRP1-C14A, and HRP1-C39A mutant protein constructs using site-directed PCR mutagenesis. Whole linear plasmid PCR products >5 kbp long were produced using Platinum® *Pfx* polymerase (Invitrogen), with the pProEX-HRP1 plasmid as template. Mutagenic primers (listed in Table 4) that overlapped the codon for the targeted amino acid, altering it to encode Ala instead of Cys, were designed. After digestion with DpnI (Invitrogen) to remove the template plasmid, the PCR products were blunt-end-ligated to form the respective full-length mutated pProEX plasmids.

Protein expression and purification

Following sequence confirmation, all five constructs were transformed into *E. coli* BL21(DE3) cells. Expression cultures were grown in LB medium to an OD_{600} of 0.6–0.7 at 37 °C, then induced with 1 mM IPTG and grown overnight (16–23 h) for full-length HRP1, HRP1-Δ128–143, and HRP1-C136A, or for 6 h for the less stable HRP1-C14A and HRP1-C39A. Cells were harvested by centrifugation and lysed on ice by sonication in 20 mM Hepes (pH 7.5), 150 mM NaCl, and 10 mM imidazole containing 1 mg of RNase A, 0.2 mg of DNase I, 10 mg of lysozyme, and one Complete Mini protease inhibitor cocktail tablet (Roche) per liter of expression culture. Selenomethionine-substituted HRP1-Δ128–143 protein was expressed in M9 minimal medium supplemented with Lys, Val, Leu, Ile, Phe, Thr, and SeMet amino acids.

For full-length HRP1, cell lysate was centrifuged to remove the insoluble material, passed through a 0.22-μm filter, and loaded onto a 5-ml HiTrap™ Chelating HP column (Amersham Biosciences) charged with Ni^{2+} and equilibrated in buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol. The protein was eluted using an imidazole gradient from 10 to 500 mM over 10 column volumes, and fractions containing the protein were pooled. rTEV protease was then added to cleave the His tag, and the mixture was left to dialyze into buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, and 5 mM β-mercaptoethanol overnight at room temperature. This cleavage left five (GAMGS) or six (GIDPFT) extra residues on the N-termini of the pProEX- and pET151-expressed proteins, respectively. The protein–protease mix was then reloaded onto a HiTrap™ column, and an imidazole gradient was run through the column again, in order to separate the untagged protein from the rTEV, cleaved His tag, and uncleaved HRP1 protein. Some of the cleaved protein eluted immediately from the column as flow-through as expected, and some of the cleaved protein bound to the

Table 4. Primer sequences

Construct name	Forward primer sequence (5′ to 3′) (mutated codon underlined)	Reverse primer sequence (5′ to 3′)	Mutation
HRP1-C136A	gcctcgcccatggccctcgccagctag[a]	gattgccttgacgaactgcacaat	Codon 136 from tgc (C) to gcc (A)
HRP1-C14A	gctgttggcgaacacgagacg	ggtcacacctgcgttcatgatgtc[a]	Codon 14 from tgt (C) to gct (A)
HRP1-C39A	gccggggacgacgaccggct	gatcggcaacgcgccgatgtc[a]	Codon 39 from tgc (C) to gcc (A)
HRP1	aggaggggatccatgaccaccgcacgcgacatcat	gtacctccgatgctcgaggataaag	
HRP1-Δ127–143	caccatgaccaccgcacgcgacatc	gacgaactgcacttaggcgtgct	

[a] Primer 5′-phosphorylated.

column and eluted during the imidazole gradient. This intrinsic metal-binding ability of the protein is consistent with the two Ni^{2+} observed bound to the structure.

Purification of the other proteins proceeded as for the full-length protein, except for native HRP1-Δ128–143, which was additionally purified using a Superdex™ 75 10/30 size-exclusion chromatography column (Amersham Biosciences) in buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, and 5 mM β-mercaptoethanol after cleavage with rTEV.

All proteins were then dialyzed into buffer without imidazole (20 mM Hepes pH 7.5, 150 mM NaCl, and 5 mM β-mercaptoethanol). The cleaved native and SeMet HRP1-Δ128–143 were concentrated (to 11 and 22 mg/mL, respectively) in preparation for crystallization trials.

Solution sizes for all proteins were estimated using analytical size-exclusion chromatography using a Superdex™ 75 10/30 size-exclusion chromatography column (Amersham Biosciences) in buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, and 5 mM β-mercaptoethanol.

Protein crystallization

All crystallization trials were carried out at 18 °C in 96-well Intelli-Plate or CrystalClear Strip plates (Hampton Research), using sitting drop vapor diffusion. Although several conditions yielded crystals, the best diffraction-quality 'cat's-eye'-shaped crystals of native HRP1-Δ128–143 grew by sitting drop vapor diffusion against a mother liquor composed of 21% methoxypolyethylene glycol 5000 and 0.2 M 4-morpholineethanesulfonic acid/KOH (pH 6.1). Crystals of SeMet substituted HRP1-Δ128–143 grew in several different conditions, but the best crystals grew as clusters of plates in 1.7 M ammonium sulfate and 0.05 M boric acid/KOH (pH 9.0).

X-ray diffraction data collection and processing

Native and SeMet-substituted HRP1-Δ128–143 crystals were flash-cooled in their respective mother liquors containing 15% and 20% glycerol, respectively, as cryoprotectant. Diffraction data from the native HRP1-Δ128–143 crystal were collected on a Rigaku RU-H3R rotating anode source operated at 50 kV and 100 mA using a MAR345 image plate detector. Data from crystals of SeMet-substituted HRP1-Δ128–143 were collected at the peak wavelength (0.9793 Å) on beamline 5.0.2 at the Advanced Light Source (Berkeley, CA) with a CCD ADSCQ315 detector. In order to collect a complete set of Friedel pairs and to improve data quality, the data were collected using a 30° slice reverse-beam protocol in the order of peak, inflection point, and remote. This involved translation of the crystal after each 30° slice of data collection. Thus, data were collected from φ to φ+30°, then from φ+180° to φ+210°, from φ+30° to φ+60°, from φ+210° to φ+240°, and so on, continuing until a full 360° of data had been collected. All data were indexed, integrated, scaled, and merged using the HKL program suite.[50]

Structure determination and refinement

The SHELX program suite[51] was used to locate the positions of the selenium atoms and to obtain an initial estimate of phases using single-wavelength anomalous diffraction with the SeMet peak wavelength data set. RESOLVE[52] was used for further density modification and phase improvement, and automatic model building was carried out with ARP/wARP.[53] Building was finalized in O,[54] and the structure was refined against peak wavelength data with CNS.[55] The partially refined model was used as the search model to solve the structure of the native protein by molecular replacement using AMoRe.[56] The model of the native structure was also completed using O and refined against the native data set using CNS.

Chaperone activity assay

Investigation of possible chaperone activity was carried out by observing whether HRP1 could prevent thermal denaturation of equine liver ADH (Sigma-Aldrich). This was performed as previously described[41] by scanning for protein aggregation using a Hitachi Fluorescence Spectrophotometer. Bovine eye lens α-crystallin (Sigma-Aldrich) was used as positive control chaperone. For each scan, solutions were mixed in a 0.5-mm quartz cuvette at room temperature with standard protein purification buffer, and a magnetic stirrer was placed in the cuvette. When pipetting each protein solution into the cuvette, the outside of the pipette tip was wiped with a clean lint-free tissue before expelling the solution. The cuvette was then placed in the spectrophotometer cuvette holder (kept at 48 °C), and a 60-min time scan started immediately with stirring. With both excitation and emission wavelengths set to 360 nm, and with excitation and emission slit widths set to 5 and 10 nm, respectively, light scattering was recorded every second throughout the scan. Between each scan, the cuvette was thoroughly rinsed with 10% bleach, tap water, and ultrapure water.

AMP binding detection using intrinsic protein fluorescence

HRP1 and PAE2307 (positive control) were diluted to concentrations of 10 and 2 μM, respectively, with various concentrations of AMP in buffer (20 mM Hepes pH 7.5, 150 mM NaCl, and 5 mM 2-mercaptoethanol). The samples were incubated for 1 min at room temperature to achieve equilibrium, and fluorescence measurements were carried out in a 5-mm quartz cuvette in a Hitachi Fluorescence Spectrophotometer at room temperature. Fluorescence was measured at the excitation and emission wavelengths producing maximum emission for each protein: 277 and 304 nm, respectively, for HRP1, and 280 and 347 nm, respectively, for PAE2307. The slit widths were set to 5 nm for both excitation and emission for all proteins. The background intensity of the buffer was subtracted from all fluorescence intensities. Changes in protein fluorescence due to the absorption of AMP (the inner filter effect) were corrected for using the following formula:[43]

$$C = 10^{\left(\frac{\Delta A_{ex}}{2} + \frac{\Delta A_{em}}{2}\right)}$$

where C is the correlation factor, and ΔA_{ex} and ΔA_{em} are the absorbances of the added titrated ligand at the excitation and emission wavelengths, respectively. This is the simplest way of correcting for this effect and assumes that fluorescence emanates from a central point in a spectroscopic cell.[43] To carry out this correction, the observed fluorescence signal was multiplied by the correlation factor C to obtain the corrected signal. Absorbance measurements were made with a Helios γ spectrophotometer (ThermoSpectronic) in the same 5-mm quartz cuvette used for fluorescence measurements.

Electron microscopy

Wild-type *M. tuberculosis* H37Rv was grown in Middlebrook 7H9 medium supplemented with oleic acid–albumin–dextrose–catalase and 0.05% (vol/vol) Tween-80. Cultures were grown with rolling until log phase (OD = 0.4–0.5 nm). The *dosR* regulon, including HRP1, was induced by standing the cultures for 24 h. This method of induction of the *dosR* regulon has been used by us and others to study the *dosR*-controlled genes.[3,57,58] Cells were pelleted by centrifugation (4000 rpm, 10 min), washed three times in sterile phosphate-buffered saline (PBS), and fixed for 24 h in 4% (vol/vol) glutaraldehyde in 0.05 M cacodylate buffer (pH 7.2). Cells were then pelleted by centrifugation as previously described and prepared for electron microscopy. Ultrathin sections (70–100 nm thick) were cut and collected on Parlodion-coated (200-mesh) nickel grids. The sections were preincubated for 15 min at room temperature with PBS [50 mM sodium phosphate buffer (pH 7.2) and 0.15 M NaCl] containing 1% (wt/vol) bovine serum albumin (BSA) to block any nonspecific binding. The sections were then incubated for 1 h at room temperature with polyclonal rabbit anti-HRP1 at 1:50 dilution in PBS/BSA. After the sections had been washed several times with PBS/BSA, they were incubated for 1 h at room temperature with a secondary antibody conjugated to 15 nm of colloidal gold (Electron Microscopy Sciences, Fort Washington, PA) at 1:25 dilution in PBS/BSA. The sections were then washed several times with water, stained with uranyl acetate (5% wt/vol in 50% ethanol) for 20 min, and washed again (once with 50% ethanol for 30 s, then twice with water at 2 min per wash). Sections were then stained with a solution of lead citrate for 2 min and washed several times with water. After the sections had been dried, they were examined with a transmission electron microscope (either a Philips Tecnai 12 electron microscope operating at 120 kV or a Philips CM12 electron microscope operating at 100 kV).

Accession codes

The refined coordinates and structure factors for both structures have been deposited at the PDB (PDB codes: native HRP1-Δ128–143, 1XKF; SeMet HRP1-Δ128–143, 1Y5H).

Acknowledgements

We thank Dr. Adrian Turner for electron microscopy services, Mr. Martin Middleditch for mass spectrometry services, and Professor Neil Stoker for helpful comments and discussions. This work was funded by the Health Research Council of New Zealand.

References

1. Wayne, L. G. & Sohaskey, C. D. (2001). Nonreplicating persistence of *Mycobacterium tuberculosis*. *Annu. Rev. Microbiol.* **55**, 139–163.
2. Park, H.-D., Guinn, K. M., Harrell, M. I., Liao, R., Voskuil, M. I., Tompa, M. *et al.* (2003). Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol. Microbiol.* **48**, 833–843.
3. Kendall, S. L., Movahedzadeh, F., Rison, S. C. G., Wernisch, L., Parish, T., Duncan, K. *et al.* (2004). The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. *Tuberculosis*, **84**, 247–255.
4. Kumar, A., Toledo, J. C., Patel, R. P., Lancaster, J. R., Jr. & Steyn, A. J. C. (2007). *Mycobacterium tuberculosis* DosS is a redox sensor and DosT is a hypoxia sensor. *Proc. Natl Acad. Sci. U. S. A.* **104**, 11568–11573.
5. Voskuil, M. I., Visconti, K. C. & Schoolnik, G. K. (2004). *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis*, **84**, 218–227.
6. Cox, R. A. (2007). A scheme for the analysis of microarray measurements based on a quantitative theoretical framework for bacterial cell growth: application to studies of *Mycobacterium tuberculosis*. *Microbiology*, **153**, 3337–3349.
7. Sharpe, M. L., Baker, E. N. & Lott, J. S. (2005). Crystallization of a protein using dehydration without a precipitant. *Acta Crystallogr. Sect. F*, **61**, 565–568.
8. Sherman, D. R., Voskuil, M. & Schnappinger, D. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding α-crystallin. *Proc. Natl Acad. Sci. U. S. A.* **98**, 7534–7539.
9. Muttucumaru, D. G. N., Roberts, G., Hinds, J., Stabler, R. A. & Parish, T. (2004). Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. *Tuberculosis*, **84**, 239–246.
10. Bacon, J., James, B. W., Wernisch, L., Williams, A., Morley, K. A., Hatch, G. J. *et al.* (2004). The influence of reduced oxygen availability on pathogenicity and gene expression in *Mycobacterium tuberculosis*. *Tuberculosis*, **84**, 205–217.
11. Voskuil, M. I., Schnappinger, D., Visconti, K. C., Harrell, M. I., Dolganov, G. M., Sherman, D. R. & Schoolnik, G. K. (2003). Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. *J. Exp. Med.* **198**, 705–713.
12. Rosenkrands, I., Slayden, R. A., Crawford, J., Aagaard, C., Barry, C. E., III & Anderson, P. (2002). Hypoxic response of *Mycobacterium tuberculosis* studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J. Bacteriol.* **184**, 3485–3491.
13. Karakousis, P. C., Yoshimatsu, T., Lamichhane, G., Woolwine, S. C., Nuermberger, E. L., Grosset, J. & Bishai, W. R. (2004). Dormancy phenotype displayed by extracellular *Mycobacterium tuberculosis* within artificial granulomas in mice. *J. Exp. Med.* **200**, 647–657.
14. Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M. *et al.* (2003). Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J. Exp. Med.* **198**, 693–704.
15. Shi, L., Jung, Y.-J., Tyagi, S., Gennaro, M. L. & North, R. J. (2003). Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. *Proc. Natl Acad. Sci. U. S. A.* **100**, 241–246.
16. Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.* **43**, 717–731.

17. Hampshire, T., Soneji, S., Bacon, J., James, B. W., Hinds, J., Laing, K. *et al.* (2004). Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: a model for persistent organisms? *Tuberculosis*, **84**, 228–238.
18. Bateman, A. (1997). The structure of a domain common to archaebacteria and the homocystinuria disease protein. *Trends Biochem. Sci.* **22**, 12–13.
19. Janošik, M., Kery, V., Gaustadnes, M., Maclean, K. N. & Kraus, J. P. (2001). Regulation of human cystathionine β-synthase by *S*-adenosyl-L-methionine: evidence for two catalytically active conformations involving an autoinhibitory domain in the C-terminal region. *Biochemistry*, **40**, 10625–10633.
20. McLean, J. E., Hamaguchi, N., Belenky, P., Mortimer, S. E., Stanton, M. & Hedstrom, L. (2004). Inosine 5′-monophosphate dehydrogenase binds nucleic acids *in vitro* and *in vivo*. *Biochem. J.* **379**, 243–251.
21. Cheung, P. C. F., Salt, I. P., Davies, S. P., Hardie, D. G. & Carling, D. (2000). Characterization of AMP-activated protein kinase γ-subunit isoforms and their role in AMP binding. *Biochem. J.* **346**, 659–669.
22. Estévez, R., Pusch, M., Ferrer-Costa, C., Orozco, M. & Jentsch, T. J. (2004). Functional and structural conservation of CBS domains from CLC chloride channels. *J. Physiol.* **557**, 363–378.
23. Hebeisen, S., Biela, A., Giese, B., Müller-Newen, G., Hidalgo, P. & Fahlke, C. (2004). The role of the carboxyl terminus in ClC chloride channel function. *J. Biol. Chem.* **279**, 13140–13147.
24. Scott, J. W., Hawley, S. A., Green, K. A., Anis, M., Stewart, G., Scullion, G. A. *et al.* (2004). CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284.
25. Adams, J., Chen, Z.-P., van Denderen, B. J. W., Morton, C. J., Parker, M. W., Witters, L. A. *et al.* (2004). Intrasteric control of AMPK via the gamma1 subunit AMP allosteric regulatory site. *Protein Sci.* **13**, 155–165.
26. Xiao, B., Heath, R., Saiu, P., Leiper, F. C., Leone, P., Jing, C. *et al.* (2007). Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature*, **449**, 496–500.
27. Townley, R. & Shapiro, L. (2007). Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase. *Science*, **315**, 1726–1729.
28. Markovic, S. & Dutzler, R. (2007). The structure of the cytoplasmic domain of the chloride channel ClC-Ka reveals a conserved interaction interface. *Structure*, **15**, 715–725.
29. Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Horer, S. *et al.* (2007). Structure of a CBS-domain pair from the regulatory gamma1 subunit of human AMPK in complex with AMP and ZMP. *Acta Crystallogr. Sect. D*, **63**, 587–596.
30. Spies, H. S. C. & Steenkamp, D. J. (1994). Thiols of intracellular pathogens: identification of ovothiol A in *Leishmania donovani* and structural analysis of a novel thiol from *Mycobacterium bovis*. *Eur. J. Biochem.* **224**, 203–213.
31. Guilvout, I., Hardie, K. R., Sauvonnet, N. & Pugsley, A. P. (1999). Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J. Bacteriol.* **181**, 7212–7220.
32. Carugo, O. & Bordo, D. (1999). How many water molecules can be detected by protein crystallography? *Acta Crystallogr. Sect. D*, **55**, 479–483.
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
34. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.
35. Bond, C. S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*, **19**, 311–312.
36. Proudfoot, M., Sanders, S. A., Singer, A., Zhang, R., Brown, G., Binkowski, A. *et al.* (2008). Biochemical and structural characterization of a novel family of cystathionine β-synthase domain proteins fused to a Zn ribbon-like domain. *J. Mol. Biol.* **375**, 301–315.
37. Lawrence, M. & Colman, P. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
38. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D*, **60**, 2256–2268.
39. Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R. & Nureki, O. (2007). Crystal structure of the MgtE Mg^{2+} transporter. *Nature*, **448**, 1072–1075.
40. Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* **33**, W89–W93.
41. Yuan, Y., Crane, D. D. & Barry, C. E., III (1996). Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: function of the mycobacterial α-crystallin homolog. *J. Bacteriol.* **178**, 4484–4492.
42. Lott, J. S., Paget, B., Johnston, J. M., Delbaere, L. T. J., Simon-Sigrell, J. A., Banfield, M. J. & Baker, E. N. (2006). The structure of an ancient conserved domain establishes a structural basis for stable histidine phosphorylation and identifies a new family of adenosine specific kinases. *J. Biol. Chem.* **281**, 22131–22141.
43. Eftink, M. R. (1997). Fluorescence methods for studying equilibrium macromolecule–ligand interactions. *Methods Enzymol.* **278**, 221–257.
44. Cunningham, A. F. & Spreadbury, C. L. (1998). Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton α-crystallin homolog. *J. Bacteriol.* **180**, 801–808.
45. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.
46. Thornton, J. M. (1981). Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261–287.
47. Abdallah, A. M., Gey van Pittius, N. C., DiGiuseppe Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E. *et al.* (2007). Type VII secretion—mycobacteria show the way. *Nat. Rev. Microbiol.* **5**, 883–891.
48. Lin, M. Y., Geluk, A., Smith, S. G., Stewart, A. L., Friggen, A. H., Franken, K. L. M. C. *et al.* (2007). Lack of immune responses to *Mycobacterium tuberculosis* DosR regulon proteins following *Mycobacterium bovis* BCG vaccination. *Infect. Immun.* **75**, 3523–3530.
49. Ganguly, N., Giang, P. H., Gupta, C., Basu, S. K., Siddiqui, I., Salunke, D. M. & Sharma, P. (2007). *Mycobacterium tuberculosis* secretory proteins CFP-10, ESAT-6 and the CFP10:ESAT6 complex inhibit lipopolysaccharide-induced NF-κB transactivation by

downregulation of reactive oxidative species (ROS) production. *Immunol. Cell Biol.* **86**, 98–106.
50. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, 267.
51. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. Sect. D*, **58**, 1772–1779.
52. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. Sect. D*, **56**, 965–972.
53. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. Sect. D*, 53.
54. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. Sect. A*, **47**, 110–119.
55. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
56. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallogr. Sect. A*, **50**, 157–163.
57. Florczyk, M. A., McCue, L. A., Purkayastha, A., Currenti, E., Wolin, M. J. & McDonough, K. A. (2003). A family of acr-coregulated *Mycobacterium tuberculosis* genes shares a common motif and requires Rv3133c (dosR or devR) for expression. *Infect. Immun.* **71**, 5332–5343.
58. Parish, T., Smith, D. A., Kendall, S., Casali, N., Bancroft, G. J. & Stoker, N. G. (2003). Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect. Immun.* **71**, 1134–1140.

Note added in proof. An alternative proposal is that both DosS and DosT are direct oxygen sensors: Sousa, E. H. S., Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. (2007). *Protein Sci*. **16**, 1708–1719.