Tuberculosis (2003) **83**, 223–249

ELSEVIER

Tuberculosis

www.elsevierhealth.com/journals/tube

# The TB structural genomics consortium: a resource for *Mycobacterium tuberculosis* biology

**T.C. Terwilliger[a,*], M.S. Park[a], G.S. Waldo[a], J. Berendzen[b], L.-W. Hung[b], C.-Y. Kim[a], C.V. Smith[c], J.C. Sacchettini[c], M. Bellinzoni[d], R. Bossi[d], E. De Rossi[d], A. Mattevi[d], A. Milano[d], G. Riccardi[d], M. Rizzi[d], M.M. Roberts[e], A.R. Coker[f], G. Fossati[g], P. Mascagni[g], A.R.M. Coates[e], S.P. Wood[f], C.W. Goulding[h], M.I. Apostol[h], D.H. Anderson[h], H.S. Gill[h], D.S. Eisenberg[h], B. Taneja[i], S. Mande[i], E. Pohl[j], V. Lamzin[j], P. Tucker[j], M. Wilmanns[j], C. Colovos[j], W. Meyer-Klaucke[j], A.W. Munro[k], K.J. McLean[k], K.R. Marshall[k], D. Leys[k], J.K. Yang[l], H.-J. Yoon[l], B.I. Lee[l], M.G. Lee[l], J.E. Kwak[l], B.W. Han[l], J.Y. Lee[l], S.-H. Baek[l], S.W. Suh[l], M.M. Komen[m], V.L. Arcus[m], E.N. Baker[m], J.S. Lott[m], W. Jacobs Jr.[n], T. Alber[o], B. Rupp[p]**

[a]*Los Alamos National Laboratory, Bioscience Division, Mail Stop M888, Los Alamos, NM, 87545, USA*
[b]*Los Alamos National Laboratory, Physics Division, Mail Stop D454, Los Alamos, NM 87545, USA*
[c]*Department of Biochemistry and Biophysis, Texas A&M University, College Station, TX 77843-2128, USA*
[d]*Department of Genetics and Microbiology, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy*
[e]*Medical Microbiology, Department of Cellular and Molecular Medicine, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK*
[f]*Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK*
[g]*Italfarmaco Research Center, Via dei Lavoratori 54, 20092 Cinisello Balsamo, Milan, Italy*
[h]*UCLA-DOE Lab of Structural Biology, Howard Hughes Medical Institute, Molecular Biology Institute, University of California, UCLA Box 951570, Los Angeles, CA 90095-1570, USA*
[i]*Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad-500076, India*
[j]*EMBL Hamburg, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany*
[k]*Department of Biochemistry, University of Leicester, The Adrian Building, University Road, Leicester LE1 7RH, UK*
[l]*Structural Proteomics Laboratory, School of Chemistry and Molecular Engineering, Seoul National University, Seoul 151-742, South Korea*
[m]*Laboratory of Structural Biology, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand*
[n]*Albert Einstein School of Medicine, Yeshiva University, Bronx NY 10461, USA*
[o]*University of California, Berkeley, CA 94720, USA*
[p]*Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*

*Corresponding author.
*E-mail address:* terwilliger@lanl.gov (T.C. Terwilliger).

1472-9792/03/$ - see front matter 
doi:10.1016/S1472-9792(03)00051-9

**KEYWORDS**
Protein structure;
X-ray crystallography;
Structural genomics;
Mycobacterium tuberculosis;
Drug discovery

**Summary** The TB Structural Genomics Consortium is an organization devoted to encouraging, coordinating, and facilitating the determination and analysis of structures of proteins from *Mycobacterium tuberculosis*. The Consortium members hope to work together with other *M. tuberculosis* researchers to identify *M. tuberculosis* proteins for which structural information could provide important biological information, to analyze and interpret structures of *M. tuberculosis* proteins, and to work collaboratively to test ideas about *M. tuberculosis* protein function that are suggested by structure or related to structural information. This review describes the TB Structural Genomics Consortium and some of the proteins for which the Consortium is in the progress of determining three-dimensional structures.


## Abbreviations

Ag85B, Antigen 85b; Cpn10, Chaperonin-10; Dsb, disulfide oxidoreductase; FAD, flavin adenine dinucleotide; GS, glutamine synthetase; ICL, isocitrate lyase; MPD, 2-methyl-2,4-pentanediol; *M. tuberculosis*, *Mycobacterium tuberculosis*; *Mtb*, *Mycobacterium tuberculosis*; Mtcpn10, chaperonin-10 from *M. tuberculosis*; NAD, nicotinamide adenine dinucleotide; NADP, nicotinamide adenine dinucleotide phosphate; NCS, non-crystallographic symmetry; NMN, N 1-methylnicotinamide; Rv, Preface for *M. tuberculosis* gene numbering system (see http://www.sanger.org); TB, tuberculosis; TBSGC, TB Structural Genomics Consortium

## Introduction

The TB Structural Genomics Consortium is an organization devoted to encouraging, coordinating, and facilitating the determination and analysis of structures of proteins from *Mycobacterium tuberculosis*. The Consortium members hope to work together with other *M. tuberculosis* researchers to identify *M. tuberculosis* proteins for which structural information could provide important biological information, to analyze and interpret structures of *M. tuberculosis* proteins, and to work collaboratively to test ideas about *M. tuberculosis* protein function that are suggested by structure or related to structural information.

The principal purposes of this review are to inform the broader *M. tuberculosis* biology community about the TB Structural Genomics Consortium, about how the Consortium would like to share information and materials such as expression clones and proteins with the community, and about how the Consortium would like to foster collaborative efforts between its members and the rest of the *M. tuberculosis* biology community.

The first part of the review describes the TB Structural Genomics Consortium, how the Consortium would like to work with the *M. tuberculosis* biology community, and how to find out more information. The following sections describe some of the work of members of the Consortium that may be of interest to the *M. tuberculosis* biology community and is intended to suggest types of interaction that might occur between members of the Consortium and other *M. tuberculosis* biologists. These sections include structural studies on a second isocitrate lyase in *M. tuberculosis* (Smith and Sacchettini), on NAD and iron metabolism (Bellinzoni et al.), on the structure of chaperonin 10 (Roberts et al.), on *M. tuberculosis* protein structures determined by the UCLA group (Goulding et al.), on a second structure of chaperonin 10 (Taneja and Mande), on plans for structure determination at the EMBL Outstation in Hamburg (Pohl et al.), and on cytochrome P450 redox systems in *M. tuberculosis* (Munro et al.), and on structural genomics of *M. tuberculosis* at Seoul National University (Yang et al.).

## Structural genomics and *M. tuberculosis* biology

First, what is structural genomics? Structural genomics is the determination of protein structures on a genomic scale. It is a new field (see references[1–3]), inspired by recent advances in technology for structure determination and by the success of the genome projects. Protein structures (the three-dimensional arrangement of atoms in a protein) are exceptionally helpful in interpreting biochemical experiments carried out on proteins and in developing hypotheses about how they work.

Members of the TB Structural Genomics Consortium think that large-scale structure determination of proteins from *M. tuberculosis* might provide a very useful framework for future *M. tuberculosis* drug discovery. A vision inspiring many members of the project is that if structures of many of the proteins from *M. tuberculosis* were known, then as key *M. tuberculosis* genes are identified by genetic means, high-throughput screening, or other approaches, the structural information that could help speed up drug discovery would already be in place. Additionally, members of the TB Structural Genomics Consortium hope that the thousands of cloned *M. tuberculosis* genes and hundreds or even thousands of purified *M. tuberculosis* proteins that they plan to produce will be useful right away to members of the *M. tuberculosis* biology community.

The TB Structural Genomics Consortium consists largely of structural biologists, as the focus of the organization is on structure determination, but several *M. tuberculosis* geneticists, biologists, and biochemists who are actively working on making structural genomics of *M. tuberculosis* possible are also members. The membership of the Consortium is available from the Consortium's web site at http://www.tbgenomics.org; over 230 individuals from 31 organizations in 11 countries are members.

The TB Structural Genomics Consortium is organized with central facilities that carry out key tasks for the benefit of the entire consortium, along with many individual laboratories working independently but sharing methods and communicating with each other about their plans and status of their work. The central facilities carry out high-throughput cloning and expression testing, crystallization, and X-ray data collection. Individuals in the Consortium work on proteins that interest them. At the same time, the central cloning and protein production facilities at Los Alamos are cloning all the *M. tuberculosis* genes targeted by any member of the consortium and testing the expression of these genes. The results of these tests are made available to the Consortium members and to the rest of the world through the Consortium web site at http://www.tbgenomics.org. The Consortium's crystallization facility crystallizes proteins purified by individual members of the Consortium and by the protein production facility. Similarly, the X-ray data collection facility collects X-ray diffraction data on crystals from individual laboratories and from the crystallization facility.

Expression clones produced by the TB Structural Genomics Consortium facilities that have been characterized are available not only to members of the Consortium, but also to members of the *M. tuberculosis* biology community. Expression clones and some purified *M. tuberculosis* proteins from the Consortium can be obtained through the *M. tuberculosis* resource at Colorado State University (see http://www.cvmbs.colostate.edu/microbiology/tb/top.htm). The data produced by the Consortium's facilities is publicly available through the Consortium web site at http://www.tbgenomics as well.

Members of the TB Structural Genomics Consortium had determined some 27 structures of *M. tuberculosis* proteins by the end of the first year and a half of the Consortium. A constantly updated list of *M. tuberculosis* protein structures determined by members of the Consortium can be found at the public web site at http://www.tbgenomics.org. The Consortium hopes that these and the structures to be determined will allow members of the *M. tuberculosis* biology community to interpret their genetic and biochemical studies in terms of the structures of these proteins.

There are several ways that members of the *M. tuberculosis* biology community can readily interact with members of the TB Structural Genomics Consortium, and most of them are facilitated by the public web site at http://www.tbgenomics.org. This site has links to email addresses of all members of the Consortium so that collaborations can be set up, and to each of the Consortium facilities, so that materials can be requested. It also lists the status of all proteins being worked on by the Consortium. It has a page where members of the *M. tuberculosis* community can tell the Consortium about *M. tuberculosis* proteins that are under investigation and the importance of these proteins.

The TB Structural Genomics Consortium is looking forward to working with other members of the *M. tuberculosis* biology community and welcomes interactions through the web site and through research collaborations.

## A second isocitrate lyase in *M. tuberculosis* (Clare V. Smith and James C. Sacchettini)

The glyoxylate shunt pathway has been shown to be essential for bacterial survival in the activated macrophage and for a persistent infection in *M. tuberculosis*.[4] The two enzymes of the shunt, isocitrate lyase (Rv0467-*icl*) and malate synthase (Rv1837c-*glcB*), are upregulated when the bacteria shift to using substrates generated by $\beta$-oxidation of fatty acids[5] and also during infection of macrophages.[6] The shunt catalyzes a carbon conserving

metabolic bypass in which isocitrate is converted to succinate and glyoxylate, then acetyl CoA is added to glyoxylate to form malate. This is a metabolic pathway which is believed to be important in the chronically infected lung of a TB patient, where fatty acids are thought to be the major source of carbon and energy. The glyoxylate pathway is present in most prokaryotes, lower eukaryotes and plants, but is not thought to operate in vertebrates. Therefore, the enzymes of the glyoxylate shunt are very attractive targets for the development of new anti-tuberculars active against persistent bacteria. Members of the Consortium have solved high-resolution structures of both isocitrate lyase (ICL) and malate synthase (MS)[7,8] and are actively involved in programs of structure-based drug design in the hope of developing new inhibitors against persistent bacteria.

The genome sequence of *M. tuberculosis* has revealed that there are two genes encoding isocitrate lyases. There are called *icl* and *aceA*, respectively[9] (Sanger website; http://www.sanger.ac.uk). In the sequence reported from *M. tuberculosis* H37Rv, *aceA* is a spilt open reading frame predicted to occur as a result of a frame shift mutation (designated Rv1915-*aceAa* and Rv1916-*aceAb*) and it is thought to be a pseudogene. Sequencing of another *M. tuberculosis* strain (CSU93) by TIGR (TIGR website; http://www.tigr.org) showed that this strain has an intact *aceA* encoding a protein of 766 amino acids with a molecular weight of 85 kDa. Subsequent sequencing of the *M. tuberculosis* Erdman strain (unpublished data, Smith and Sacchettini) and 210 strain (R. Fleischmann, personal communication) indicates that the same gene is present in these strains and would encode a long ICL of approximately 85 kDa.

The *aceA* gene was annotated based on its sequence homology to other known ICL enzymes. It has 23% identity to both *aceA* from *Escherichia coli* and *icl* from *M. tuberculosis* and also contains regions of conserved sequence which have been shown to be important in catalysis (Fig. 1a). For example, *aceA* contains the signature catalytic motif consisting of residues KKCGH, which have been previously identified at positions 193–197 in *E. coli* using site-directed mutagenesis.[10–13] The importance of these residues in the active site has been confirmed by the structure of ICL from *M. tuberculosis* in complex with a substrate-like analogue.[7] More convincingly, the *aceA* gene has been cloned, expressed and the purified protein shown to have isocitrate lyase activity.[5] Kinetic analysis shows that AceA has a lower $V_{max}$ and a higher $K_m$ value of 1.3 mM for the substrate isocitrate than ICL whose $K_m$ is 145 μM. AceA is also inhibited by known ICL inhibitors such as 3-nitropropionate and bromopyruvate. In addition to this, a protein has been detected by Western blots of lysates from *M. tuberculosis* CSU93 and Erdman using an antibody raised against an internal peptide of AceA, confirming its expression in the bacterium. No expression was observed in *M. tuberculosis* H37Rv, supporting the idea that *aceAa*/*aceAb* is indeed a pseudogene in this particular strain.

Many genome sequences have become available in the last couple of years and have shown other organisms to have two isocitrate lyase-like enzymes. For example, in *Saccharomyces cerevisiae* two open reading frames designated ICL1 (YER065c) and ICL2 (YPR006c) were identified (*Saccharomyces* Genome Database website, http://genomewww.stanford.edu/Saccharomyces). ICL1 encodes an isocitrate lyase which is essential for growth on ethanol and acetate, while the highly homologous ICL2 gene has also been shown to be utilized for growth on ethanol.[14] *Aspergillus nidulans* has two similarly related enzymes named *icl* and *mclA*.[15] In prokaryotes, such as *Salmonella typhimurium,* a second gene homologous to *icl* has also been identified in the *prpBCDE* operon called *prpB*.[16] In all these cases however the second ICL-like gene has now been characterized as encoding a 2-methylisocitrate lyase, a very closely related enzyme involved in the methylcitrate cycle. In the methylcitrate cycle of propionyl-coenzyme A metabolism 2-methylisocitrate is converted to succinate and pyruvate, a reaction similar to that catalyzed by isocitrate lyase. Sequence comparisons between authentic ICLs and 2-methylisocitrate lyases indicate that the catalytic motif of the 2-methylisocitrate lyases can be recognized by the sequence KRCGH, rather than KKCGH.[15] We know that AceA contains the KKCGH motif which suggests that in the case of *M. tuberculosis* we have two genuine isocitrate lyase enzymes.

Typically prokaryotes encode an ICL which is 45–50 kDa in size and lower eukaryotes such as *A. nidulans* and *Candida albicans* encode a protein of approximately 60 kDa. To date, mycobacteria are the only example with this extra long ICL. The genome sequence of *M. leprae* encodes a 798 amino acid protein of 89 kDa called AceA (ML1985).[17] From the currently available sequence *M. bovis* and *M. avium* also encode this long version ICL (Sanger website, http://www.sanger.ac.uk; TIGR website, http://www.tigr.org) and so it appears to be a mycobacteria-specific protein. Comparison of the typical prokaryotic enzyme sequence with that of lower eukaryotes indicates an insertion of approximately 100 amino acids towards the C-terminus. In AceA from *M. tuberculosis* this insertion occurs at

the same place and is about 150 amino acids. Also AceA has an extension of about 200 amino acids at its C-terminus (Fig. 1a). The prokaryotic enzymes consist of an $\alpha/\beta$ barrel and are active as tetramers.[7,18] The residues which are important for catalysis are located in this domain and sit at the C-terminal ends of the $\beta$ strands[7] (Fig. 1b). In *A. nidulans*, the insertion of 100 amino acids folds to form a five helical domain called domain II (Fig. 1c). It has been postulated that domain II might play a role in sequestration of the eukaryotic enzyme to the peroxisomal microbodies[18] but no function has yet been suggested for this domain of AceA in mycobacteria. One intriguing question about this class of mycobacteria-specific isocitrate lyases is the role of the additional 200 or so amino acids at the C-terminus. Homology searches have not yet shed light on a function for this domain, though suggestions have included a role in localization or regulation.

Experiments are currently underway to try and answer questions about the role of AceA. It seems unlikely that these two isocitrate lyases are merely redundant enzymes in TB. In fact we know that the deletion of *icl* alone cannot be compensated for by *aceA*.[4] Biochemical studies with the purified enzyme tend to suggest that AceA is a less-efficient lysis enzyme than ICL which may play a subordinate role to ICL in *M. tuberculosis* or may not be involved in the glyoxylate shunt at all. It is possible that the extra domains in AceA give the enzyme additional activities as yet unknown. Members of the TBSGC are working towards structural answers to these questions. A high-resolution structure of this unique long ICL may shed light on the roles of the extra domains by structural comparisons with other proteins in the database. More significantly in terms of treatment of TB is the possibility of structure-based drug development. Using the details of the active site of both ICL and AceA together, an inhibitor which is potent against both enzymes is a possibility, so increasing the efficacy of inhibitors and reducing the chance of multi-drug resistance development, desirable features of any new anti *M. tuberculosis* drug.

## Structural studies on NAD and iron metabolism in *M. tuberculosis* (M. Bellinzoni, R. Bossi, E. De Rossi, A. Mattevi, A. Milano, G. Riccardi and M. Rizzi)

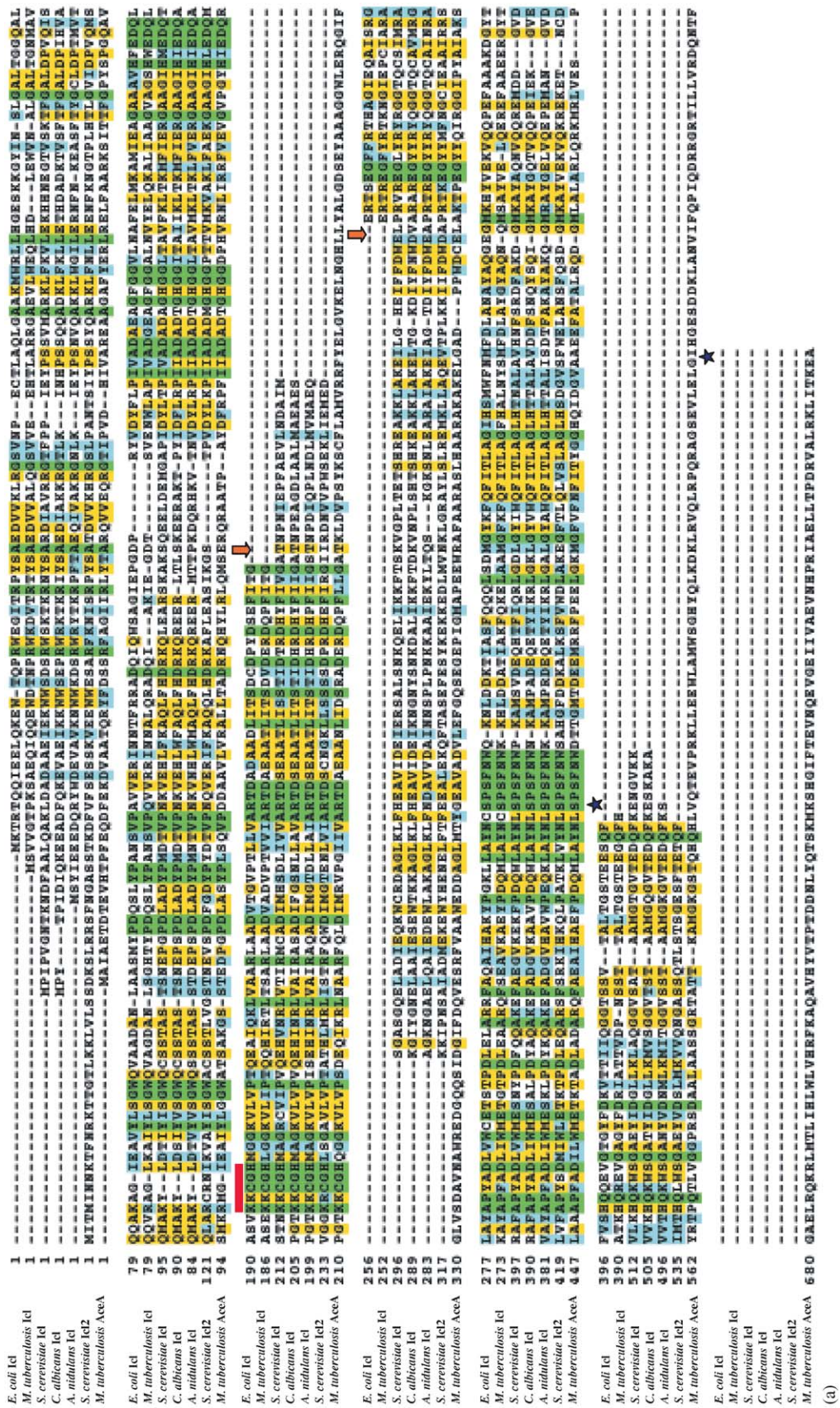
### NAD biosynthesis

Being a vital cofactor for cell survival in all living organisms, NAD metabolism must be tightly regulated in the cell. The biosynthesis of NAD shows profound differences between organisms: NAD biosynthesis proceeds through tryptophan degradation in eukaryotes while a de novo biosynthetic pathway, starting from aspartic acid, characterizes prokaryotes.[20] These two major branches merge at the level of quinolinic acid, which is then transformed into NAD by the action of two ubiquitous enzymes: NMN adenylyltransferase and NAD synthetase.[20] Once synthesized, NAD can be recycled within salvage routes.[20] However, *M. tuberculosis* appears to lack a salvage route and relies entirely on de novo biosynthesis for its NAD supply. We are presently working on several enzymes involved in NAD biosynthesis which we consider to be attractive targets for the design of new anti-tubercular drugs.

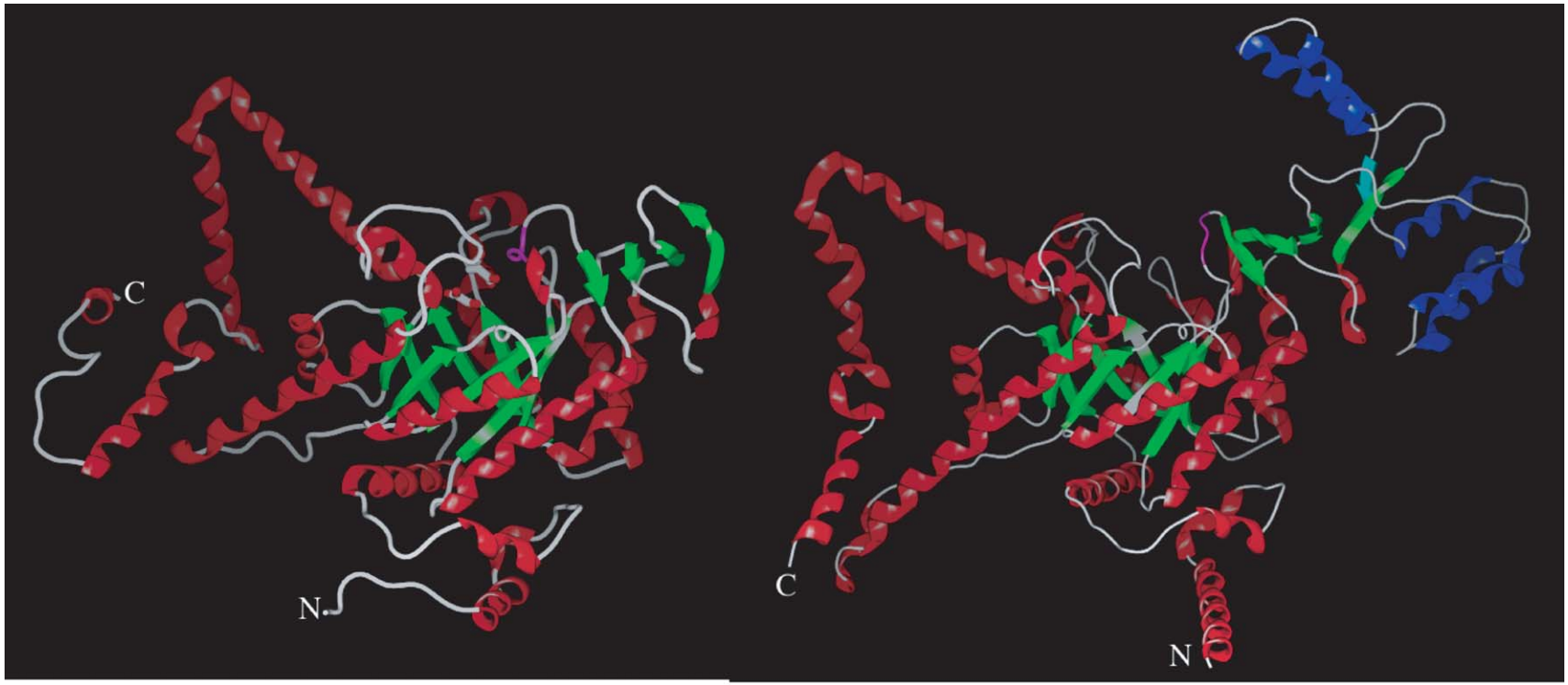
NAD synthetase belongs to the amidotransferase family and catalyses the last step in NAD biosynthesis, transforming deamido-NAD into $NAD^+$ by a two-step reaction. The only known 3D structure of this enzyme is that of the strictly ammonia-dependent *Bacillus subtilis* $NAD^+$ synthetase.[21] We demonstrated that the *M. tuberculosis* Rv2438c gene, coding for a polypeptide chain 738 residues long, is the structural gene of $NAD^+$ synthetase.[22] The enzyme shows a striking difference in the subunit theoretical dimension (80 kDa) with respect to the other bacterial enzymes characterized to date. Moreover, the *M. tuberculosis* enzyme can use both ammonia and glutamine as nitrogen source and was suggested to represent a member of a new family of amidotransferases.[22] The over-expression of *M. tuberculosis* $NAD^+$ synthetase was attempted both in *E. coli* and in cultured *Thricoplusia ni* insect cells, by means of baculovirus vector. Using a variety of vectors, bacterial strains and experimental conditions, we observed either a low level of expression in soluble form or an unsatisfactory purification yield. We then identified a stable fragment which has recently been expressed in *E. coli*. The recombinant fragment was purified and has shown a six-fold higher specific activity than the wild-type.[23] Crystallization trials of this protein are in progress.

### Iron metabolism

Iron acquisition is thought to be a key factor in the survival of intracellular pathogens such as *M. tuberculosis*, which is also strictly dependent on heme-containing enzymes for survival. We selected two potential targets for their role in iron metabolism: FprA, and the two related proteins FurA and FurB.


E. coli lcl
M. tuberculosis lcl
S. cerevisiae lcl
C. albicans lcl
A. nidulans lcl
S. cerevisiae lcl2
M. tuberculosis AceA

(a)




(b)

**Fig. 1** (a) Multiple sequence alignment of *M. tuberculosis* ICL and AceA with other isocitrate lyases. Sequences from *Escherichia coli* ICL (434 amino acids), *Mycobacterium tuberculosis* ICL (strains H37Rv and CSU93, 428 amino acids), *Saccharomyces cerevisiae* ICL (557 amino acids), *Candida albicans* ICL (550 amino acids), *Aspergillus nidulans* ICL (536 amino acids), *Saccharomyces cerevisiae* ICL2 (575 amino acids) and *Mycobacterium tuberculosis* (strain CSU93) AceA (766 amino acids) were aligned using Clustal W.[19] Conserved residues are marked in green. The catalytic motif, KKCGH, is shown by the red bar and is conserved in all known isocitrate lyase enzymes. Note this motif is KRCGH in *Saccharomyces cerevisiae* ICL2 which is a 2-methylisocitrate lyase. The insert of approximately 100 amino acids in lower eukaryotic enzymes is marked by the orange arrows. This insert is also seen in a similar position in *M. tuberculosis* AceA and is approximately 150 amino acids. The C-terminal extension of the mycobacterial AceA is marked by blue stars. (b) A ribbon representation of the three-dimensional structure of *M. tuberculosis* ICL. The $\alpha$-helices are shown in red, $\beta$-strands in green. The N- and C-termini are marked by N and C, respectively. The structure of ICL was solved in complex with glyoxylate and 3-nitropropionate, which are shown in ball and stick representation sitting at the C-terminal end of the $\beta$-sheet. The loop which contains the catalytic motif KKCGH is indicated in purple. (c) A ribbon representation of the structure of *A. nidulans* ICL. The eukaryotic enzyme contains a very similar $\alpha/\beta$ barrel to that seen in *M. tuberculosis* ICL ($\alpha$-helices are shown in red and $\beta$-strands in green). The position of the loop containing the catalytic residues is indicated in purple sitting over the top of the C-terminal ends of the $\beta$-sheet. The insertion of approximately 100 amino acids seen in the eukaryotic enzyme forms a second domain, called domain II and is shown with $\alpha$-helices in blue and $\beta$-strands in cyan.


**Fig. 2** Ribbon representation of the dimer of *M. tuberculosis* FprA. The FAD and $NADP^+$ cofactors are shown as ball-and-stick.

### FprA

*M. tuberculosis* FprA (coded by the gene Rv3106) is a 50 kDa flavoprotein whose exact role has still to be established. FprA shows significant sequence identity with the adrenodoxin reductase enzyme of mammals and with its yeast homologue. This information suggests a possible involvement of FprA either in iron metabolism or in the cytochrome P450 reductase activity. Given the major role of both processes for *M. tuberculosis* survival, studies on FprA are of interest for a better understanding of the metabolism of this organism. We have determined the structure of *M. tuberculosis* FprA at atomic resolution in its oxidized (at 1.05 Å) and reduced (at 1.25 Å) forms in complex with $NADP^+$.[24] FprA consists of a tight homodimer where each subunit contains a FAD-domain (residues 2–108 and 324–456) and an NADPH-domain (residues 109–323), both domains exhibiting the dinucleotide binding fold topology (Fig. 2). The high accuracy resulting from very high-resolution data set allows a description of both the proteins and the FAD and NADP cofactors in high detail. The active site of the fully oxidized and of the reduced structures is nearly identical, showing no relevant conformational changes. However, a careful analysis of the oxidized FprA:NADP complex reveals an unexpected and novel covalent adduct on the pyridine nucleotide. Such an unusual feature could be of use in the design of specific inhibitors.

### FurA and FurB

Fur (ferric uptake regulator) is the best-studied iron-responsive transcriptional regulator.[25] In Gram-negative and some Gram-positive bacteria, Fur is known to control the expression of the genes involved in iron uptake. When complexed to $Fe^{2+}$, its co-repressor, Fur binds to a specific palindromic consensus promoter sequence and inhibits transcriptional initiation of iron-regulated genes. Fur is a central component in the complex network involving iron metabolism, acid-tolerance response, sugar metabolism, toxin synthesis and oxidative stress response. Iron in turn is involved in the regulation of virulence factors in a wide variety of bacteria.[26]

The *M. tuberculosis* genome contains two open reading frames which are homologous to Fur in other species. These are the genes *furA* (Rv1909c) and *furB* (Rv2359) whose functions are being investigated in several laboratories. Recently, Zhart et al.[26] demonstrated that in *M. smegmatis* a knock-out *furA* mutant overexpresses the catalase-peroxidase KatG and that this altered phenotype is complemented by expression of *FurA*; thus, *FurA* appears to be a negative regulator of *katG*. Milano et al.[27] have investigated the transcriptional activation of *furA* and *katG* upon oxidative stress in *M. smegmatis*.

Completely unknown is the role of the *furB* gene, whose presence is demonstrated in *M. tuberculosis* and *M. leprae* (accession number AL583919), but not in other mycobacteria. A *furB* homolog was recently found in *M. smegmatis* (A. Milano, unpublished results). *M. tuberculosis* FurB is more similar to *E. coli* Fur and may act as a global regulator affecting gene expression in response to a variety of signals. Since FurA seems involved in oxidative stress response, FurB might be more directly involved in iron metabolism.

*M. tuberculosis* FurA and FurB proteins were overexpressed in *E. coli* and purified. At present we are beginning crystallization trials on these proteins. We are also analyzing *M. tuberculosis* (and *M. smegmatis*) FurB protein by circular dichroism to detect structural changes in response to the presence of different metal ions or other environmental conditions.

## The self-association of *M. tuberculosis* chaperonin 10 heptamers (Michael M. Roberts, Alun R. Coker, Gianluca Fossati, Paolo Mascagni, Anthony R.M. Coates and Steve P. Wood)

Chaperonin 10 (GroES) acts in concert with the 60 kDa chaperonin, GroEL, and a repertoire of other

molecular chaperones to catalyse protein folding.[28] The genome for *M. tuberculosis* contains the two GroEL homologues Mtcpn60.1 and Mtcpn60.2[29] that presumably function in concert with Mtcpn10 to enable protein folding in the mycobacterium to occur and to contribute to the stress response of the organism that resides in the hostile interior of the macrophage. We have previously reported Mtcpn10 crystallization conditions, data collection, molecular replacement (MR) solution with *E. coli* GroES and initial rigid body refinement.[30] At that stage, the appearance of new electron density clearly showed that two heptamers are linked as a tetradecamer by interleaving base loop interactions that extend between them.[30] More recently, a lower resolution model (3.5 Å) for the Mtcpn10 structure was reported by Taneja and Mande[31,32] that is in disagreement with our model. Electron density at the ends of the base loops were absent, making their loop connectivity ambiguous. Here we describe our completed Mtcpn10 structure and compare it with that of Taneja and Mande. We conclude here that Taneja and Mande have been investigating the same molecular assembly, which they have interpreted differently.

The Mtcpn10 model was refined by simulated annealing,[33] torsion angle[34] and positional refinement in the programmes *X-PLOR*[35] and *CNS*[36] between rounds of model-building with the programmes *QUANTA97* (MSI) and *SwissPdbViewer*,[37] yielding the refinement statistics listed in Table 1. To improve the electron density maps, NCS averaging and phase extension from 8.0 to 2.8 Å was carried out with the programme *DM* in the *CCP4* suite.[38] To avoid NCS bias, 5% of the data was reserved to form an $R_{free}$ test set, selected from a series of thin resolution shells using *DATAMAN*.[39] All data from 30–2.8 Å were used in the refinement, calculation of electron density maps and solvent addition to give $R_{free}$ (0.280) and $R_{work}$ (0.259) for the final model. The geometry for the final model is better than that for an average protein structure determined at this resolution (Table 1). Buried molecular surface calculations used the algorithm of Lee and Richards[40] in *CNS* with a probe radius of 1.4 Å.

The crystal structures of three other chaperonin 10 proteins are known,[41–43] and all have subunit folds similar to Mtcpn10. Mtcpn10 is a tetradecamer with a spherical cage-like structure composed of two copies of the dome-shaped heptamer each with a hydrophilic inner cavity and loops extending outwards from the base (Fig. 3a). These base loops are referred to as mobile loops in cpn10 structures because they are typically disordered.[41,42] In the *M. leprae* cpn10[42] and T4 phage Gp31[43] structures the base loops also make contact with their symmetry-related heptamers, although the inter-heptamer contacts are not as extensive as those in Mtcpn10. The two Mtcpn10 heptamers interact through their base loops, which become visible as a result of their immobilization on binding (Fig. 3a). The base loops have a significant area of contact with the subunits of the opposite heptamer that is slightly less than the subunit interactions within the heptamer. Both heptamers are positioned on the

**Table 1** Data collection and refinement statistics.

| |
|---|
| Data collection |
| Unit-cell parameters at 100 K: 76.5 × 87.9 × 124.4 Å (90°, 106.8°, 90°) |
| Resolution: 30.0–2.8 Å; no. of measurements: 134,674 |
| No. of unique reflections: 38,321; multiplicity: 3.5 |
| Completeness: 98.0%; $R_{sym}$[a]: 5.0% |
| Refinement |
| Target: maximum likelihood function |
| Anisotropic B-factor correction: applied to 6.0–2.8 Å data |
| Bulk solvent: correction applied with overall density level of 0.26 eÅ$^{-3}$ |
| No. of reflections (working/test): 34,993/1900 with $F_{obs} > 0$ |
| $R_{work}$[b]/$R_{free}$: 22.9/27.5% |
| Total no. of atoms: 10,690; no. of waters: 64; no. of MPD molecules: 18; no. of calcium ions: 1 |
| Bond length deviation (rms): 0.011 Å; bond angle deviation (rms): 1.398 Å |
| Dihedral angle deviation (rms): 26.1°; improper angle deviation (rms): 1.0° |
| *PROCHECK* [48] data: 99% of residues in allowed regions of the Ramachandran plot; overall *G*-factor 0.16 |
| *SFCHECK* [49] data: Luzzatti coordinate error:[50] 0.46 Å; Matthews coefficient:[51] 2.67 |
| Solvent content: 53.66%; correlation factor: 0.89 |

[a] $R_{sym} = \sum_h \sum_j |I_{hj} - \langle I_h \rangle| / \sum_h \sum_j I_{hj}$, where $I_{hj}$ is the intensity of observation $j$ of reflection $h$.
[b] $R_{work} = \sum_h ||F_o| - |F_c|| / \sum_h |F_o|$, where $F_o$ and $F_c$ are observed and calculated structure factors, respectively.

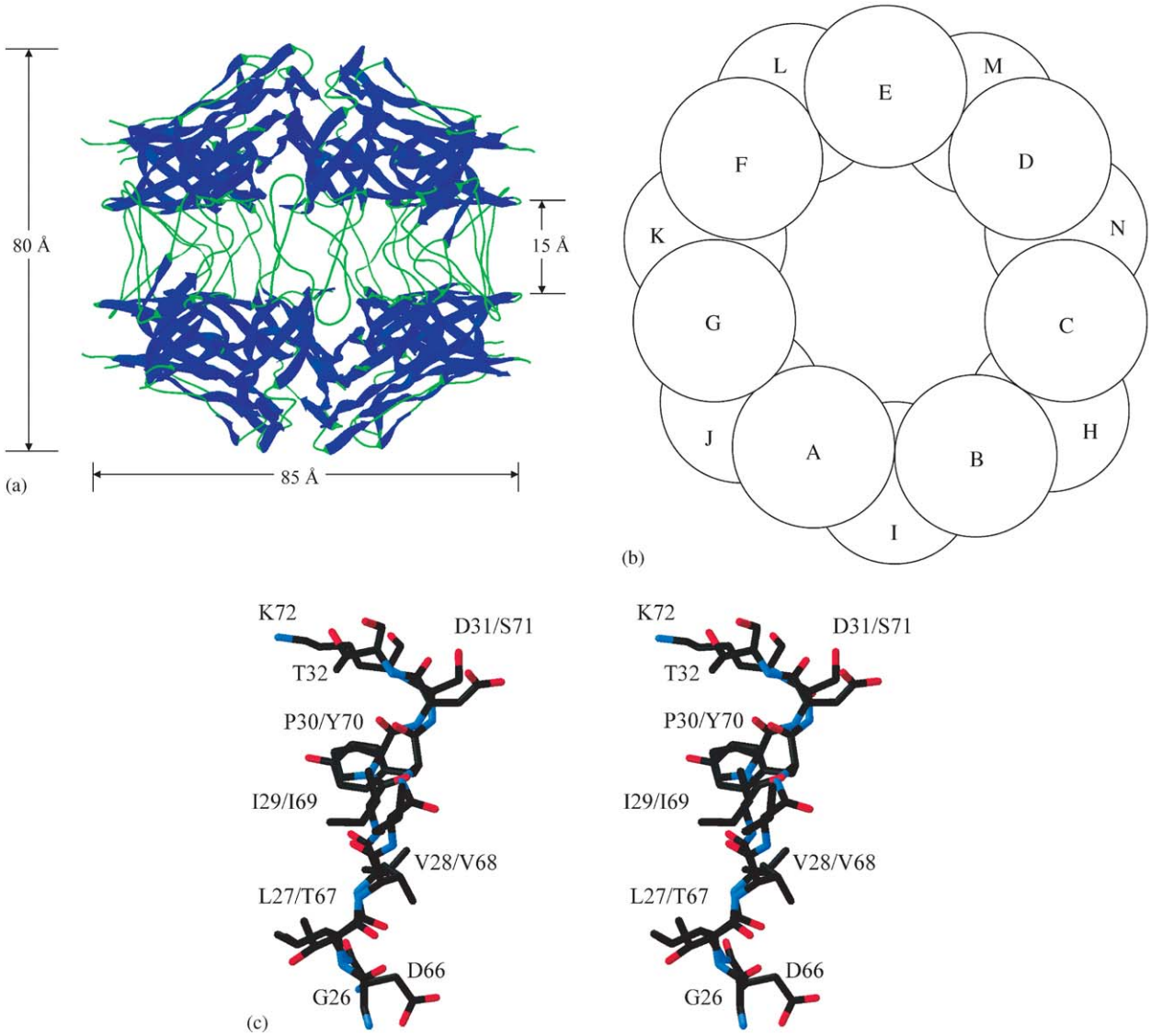



**Fig. 3** The Mtcpn10 tetradecamer. (a) Viewed along 2-fold NCS axis. (b) Sketch of the staggered arrangement of subunits viewed down the seven-fold NCS axis. (c) Superposition of polypeptide segments 26–32 and 66–72 of subunit K associated with bone resorption activity. This figure was prepared using *SwissPdbViewer*.

same seven-fold axis and related by two-fold non-crystallographic symmetry (NCS) as the tetradecamer that defines the asymmetric unit of the Mtcpn10 crystal structure.[30] This means seven NCS two-fold axes run perpendicular to the seven-fold axis in between pairs of base loops, relating subunit pairs in opposite heptamers. There is a staggered arrangement of heptamers about the seven-fold axis, with a half-subunit rotation of heptamer 1 (subunits A-G) with respect to heptamer 2 (subunits H-N) (Fig. 3b). The core domain of each subunit is a $\beta$-barrel of two orthogonal antiparallel $\beta$-sheets. The $\beta$-sheet interactions on the $\beta$-barrels extend across the subunit interfaces of the heptamer between the N- and C-terminal $\beta$-strands of neighboring subunits in a similar fashion to other cpn10 structures.[41–43] Each heptamer has a narrow oculus at the dome roof surrounded by a cluster of acidic residues which are held together by either chelated divalent metal cations or protonated sidechains at acidic pH.

The base loop (amino acids 17–33) extends out of the $\beta$-barrel domain to interact with subunits in the neighboring heptamer (Fig. 3a). The three types of subunit interface connect where the hydrophobic triplet Leu 27–Val 28–Ile 29 in the base loop of heptamer 1 forms a part of the $\beta$-sheet with heptamer 2 (or vice versa). This forms a hydrophobic patch made up of contributions from three subunits. Val 28 in the base loop of subunit I for

example, forms a $\beta$-sheet with Lys 79 of subunit A. This places Leu 27 of subunit I in a position to interact with Val 3, Ile 5 and Ile 78 of subunit A and Ile 69 and Val 97 of subunit B, reinforcing the hydrophobic interactions between subunits in the heptamer.

The base loops are essential for complexation with cpn60.[44,45] These loops are mostly disordered in uncomplexed GroES[41] and *M. leprae* cpn10.[42] This is the least-conserved sequence amongst the cpn10s and yet the five-residue segment in the middle of the loop consistently adopts a $\beta$-turn structure when it is immobilized, as shown by structures of tetradecameric Mtcpn10, Gp31[43] and the GroEL-GroES-(ADP) complex.[45] This is corroborated by NMR measurements, which show turns in the base loops of GroES and Gp31 as part of a $\beta$-hairpin structure when bound to GroEL.[46] X-ray crystallographic and NMR data indicate that the base loop of human cpn10 also preferentially samples a hairpin conformation.[47] However, there is conformational variability in the base loop sequences surrounding the $\beta$-turn in all these structures.

Another model of the Mtcpn10 structure, determined at a resolution of 3.5 Å, has been reported by Taneja and Mande.[31,32] Although it was reported as a heptamer, the two models are very similar in many respects, and have an rmsd of the superimposed C$\alpha$ coordinates of 0.97 Å (after applying the crystallographic symmetry operation $-x$, $y$, 1/2-$z$ to the coordinates of their model to generate the tetradecamer). The connectivity of the models differ, however. There are some similarities in the crystallization conditions in the two cases: sodium acetate buffer with calcium chloride in each case, but the pH was 4.0 in their case[31,32] and 5.4 in ours.[30] Consequently it seems likely that the structure of the protein in the two cases is the same, but in different crystal forms. In the structure of Taneja and Mande, electron density was absent in the base loops from residues 17–21 and from residues 31–36, leaving a disconnected loop from residues 22–30. Taneja and Mande concluded that these disconnected loops were joined to each of the seven subunits in a novel configuration, folded into the same heptamer.[31,32] This corresponds to the *cis*-configuration in the tetradecamer. In our structure, determined at higher resolution (2.8 Å compared to 3.5 Å) and showing more details, we find that residues 17–21 and 31–36 connect residues 22–30 to the subunits of the opposite heptamer in *trans* in a similar fashion to the Gp31 structure.[43] To verify our interpretation, we have calculated "simulated annealing omit" maps at the main chain junctions in question, and these show that the base loops extend towards the symmetry-related heptamer. Additional evidence for this is seen in the electron density map, based on *DM* phase extensions with the initial rigid-body-refined MR model before any model building was done. Although the MR model had gaps between residues 15–35, 48–58, and 69–73, the connectivity of the base loop is unambiguous in the electron density. There is the remote possibility that the Taneja and Mande model undergoes a "mobile loop flip" relative to ours, at the slightly lower pH, but there is no evidence for this in the electron density map.

Our earlier work showed that Mtcpn10 has bone resorption (BR) activity and that the sequences 26–35 and 65–70 were identified as being the osteolytically active regions that stimulate the growth of osteoclasts, the cells involved in BR.[102] Superimposition of these two sequences shows that they have quite similar structural features (Fig. 3c). It is therefore possible for both sequences to share the same cell receptor. The identification of the common structural features of both BR sequences should make it possible to aid BR receptor mapping and to identify pharmacophoric points for structure-based drug design against spinal tuberculosis. There is no evidence to show that calcium binding by Mtcpn10 has any role in BR.[32] It should also be stressed that the BR sequences do not form a single conformational unit[32] but are located on quite separate parts of the Mtcpn10 structure.

## Recent structures of *M. tuberculosis* proteins and their biological relevance (Celia W. Goulding, Marcin I. Apostol, Daniel H. Anderson, Harindarpal S. Gill and David Eisenberg)

The rise in TB incidence over the two last decades is due partly to TB deaths in HIV-infected patients and partly to the emergence of multidrug resistant strains of the bacteria. This rapid increase in the disease has led to the World Health Organization funding a large effort towards controlling this disease before it becomes a global epidemic.[52] In the past four decades, molecular understanding of the specialized pathogenic strategies of *M. tuberculosis* has greatly increased due to the development of efficient mutagenesis techniques to test for essential genes,[53] and the completion and annotation of the *M. tuberculosis* genome at the Sanger Center.[9] We aim to determine *M. tuberculosis* protein structures which are potential drug targets and also will contribute to greater

understanding of growth, persistence and virulence of *M. tuberculosis*.[54] In this review we will describe the *M. tuberculosis* protein structures solved at UCLA and their biological relevance for *Mtb*.

Two of the five UCLA *M. tuberculosis* protein structures were solved by multiwavelength anomalous diffraction from selenomethionine-containing protein derivatives (Rv1926c and Rv2878c), the remaining three (Rv1886c, Rv2220 and Rv3607c) were solved by molecular replacement. Purification of each protein and methods for structure determination have been described previously.[55]

## Glutamine synthetase—Rv2220

Glutamine synthetase (GS) is encoded by the *glnA1* gene from *Mtb*, and has been observed extracellularly during the early stages of infection.[56,57] *M. tuberculosis* GS may play a role in the synthesis of poly (L-glutamine-L-glutamate) chains,[58] a constituent unique to pathogenic mycobacterial cell walls, comprising 10% of bacterial mass.[59] These chains have been reported to be tightly associated with the peptidoglycan layer of the cell wall and therefore may play an important role in *M. tuberculosis* virulence. In parallel with the structure solution of *M. tuberculosis* GS,[60,61] the laboratory of Marcus Horwitz at UCLA has demonstrated that two known bacterial GS inhibitors, L-methionine-*S*-sulfoximine and phosphinothricin, selectively block the growth or normal cell wall development of *M. tuberculosis* by inhibiting extracellular GS molecules,[56] suggesting that *M. tuberculosis* GS is an attractive drug target.

In our laboratory, we have solved the X-ray crystallographic structures of *M. tuberculosis* GS[59,60] and *Salmonella typhimurium* GS.[62] Each *glnA* gene isolated from both bacilli were separately sub-cloned into an *E. coli* expression vector; the functionality of the products were demonstrated by their ability to restore growth of an *E. coli* glutamine auxotroph, having similar growth rates. Crystals of the recombinant *M. tuberculosis* GS were found, and X-ray diffraction data were collected to 2.4 Å resolution.[60] The structure of *M. tuberculosis* GS elucidates the mechanism of regulation of bacterial GS and will be described in detail elsewhere.[61] The crystal packing is shown in Fig. 4. In addition, we have co-crystallized *S. typhimurium* GS, which shares a 52% identity with TB-GS, with phosphinothricin. Together, both GS models can be used for future work in computational drug-design studies.

Interestingly, GS is observed extracellularly around pathogenic mycobacteria such as *M. tuberculosis* and *M. bovis*, but is not secreted by nonpathogenic mycobacteria such as *M. smegmatis*, thereby implicating GS in pathogenectiy of *Mtb*. It is unclear how GS is released by pathogenic mycobacteria. It has been shown that recombinant *M. tuberculosis* GS expressed in *M. smegmatis* is released into the extracellular region.[63] This suggests that *M. smegmatis* potentially contains the appropriate secretion mechanism, but its endogenous GS enzyme lacks information which is inherent in the *M. tuberculosis* homologues for targeting to the appropriate pathway. *M. tuberculosis* GS lacks a leader sequence, implying that the critical information dealing with its export is either within the amino acid sequence or a feature of bacterial leakage/auto-lysing.[64] Unfortunately the structure of *M. tuberculosis* GS revealed no additional information on this mechanism.

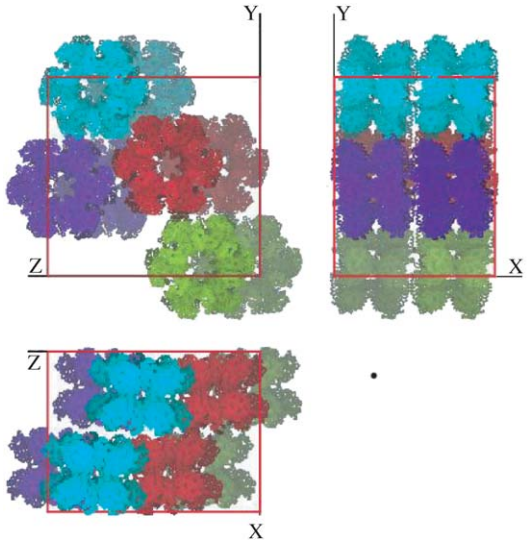



**Fig. 4** Three-dimensional structure of *M. tuberculosis* glutamine synthetase. The figure shows the crystal packing of *M. tuberculosis* GS molecules in the $P2_12_12_1$ crystal form. Three orthogonal views of the unit cell are illustrated. The action of the four symmetry operators are represented by different colors. Two dodecamers of identical color comprise one asymmetric unit.

## Ag85B—Rv1886c

*Mtb* expresses three closely related mycolyl tranferases, also known as antigen 85 proteins [ag85A, ag85C (32 kDa each); ag85B (30 kDa)]. All three contribute to cell wall synthesis by catalyzing transfer of mycolic acid from one trehalose 6-monomycolate to another, resulting in trehalose 6,6′-dimycolate and free trehalose.[65] Mycolic acids are long-chain, $\alpha$-alkyl-$\beta$-hydroxyl fatty acids that

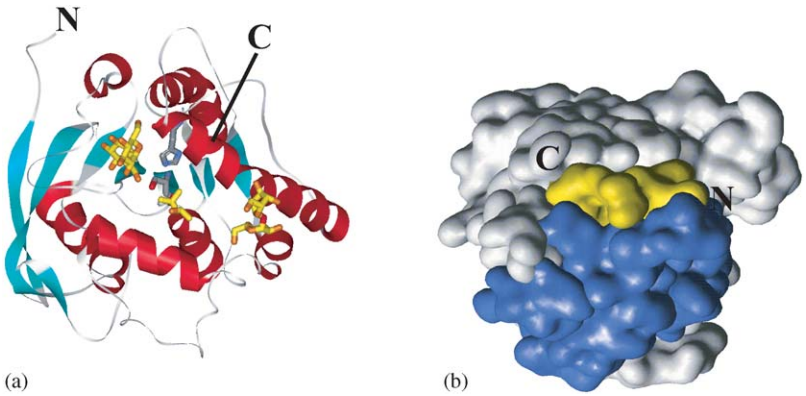



**Fig. 5** Three-dimensional structure of *M. tuberculosis* antigen 85B. (a) Ribbon diagram of trehalose-bound ag85B, showing its $\alpha/\beta$-hydrolase fold. The structure is color coded by secondary structure. Red ribbons represent $\alpha$-helices, blue ribbons $\beta$-strands, and coil and turn regions are white. The N and C termini are indicated in the illustration. Active site Ser126 and His262 C$\alpha$ and side-chain atoms are shown (protein C atoms are gray; N, blue; O, red). Two trehalose and the central MPD molecules are shown with yellow for C atoms and orange for 0 atoms. This figure was produced using XtalView, and were rendered with Raster3D. (b) Fibronectin binding region of ag85B. The diagram depicts the surface of ag85B. The yellow area represents the residues shown to be involved in fibronectin binding, and the blue regions represent the other regions involved in fibronectin binding.

form a major component of the *M. tuberculosis* cell wall. The 30 kDa major secreted protein (ag85B) is the most abundant protein exported by *Mtb*, and is a potent immunoprotective antigen.[66] Due to its location at the cell wall and its involvement in cell wall biogenesis, the ag85B protein may be a relatively accessible drug target. Inhibition of ag85B may block the synthesis of the cell wall and therefore inhibit growth of *Mtb*. Mycolyl transferase activity is unique to mycobacteria, implying that there is little chance that an inhibitor of ag85B would also inhibit a human enzyme.

The structure of ag85C has been previously solved,[67] and has 73% amino acid sequence identity with the ag85B. The ag85B and ag85C structures are similar. The structure of ag85B[68] consists of a single-domain monomeric protein that adopts an $\alpha/\beta$-hydrolase fold,[64] Fig. 5a. Interestingly, the structure contained two bound trehalose molecules and a 2-methyl-2,4-pentanediol (MPD) molecule. It has been proposed that MPD could mimic the partially hydrophilic head of the intermediate mycolate ester. Close to these sites is a hydrophobic groove that could be the binding site of the mycolate $\beta$-chain. In addition to its importance in cell wall biosynthesis, ag85 proteins stimulate the uptake of mycobacteria bacilli by human macrophages. Ag85 proteins have been proposed to interact with the gelatin binding site of human fibronectin, resulting in the enhancement of complement-mediated phagocytosis by host macrophages. The region which is responsible for fibronectin-binding is approximately 29% of accessible surface area of the protein,[63,69] Fig. 5b. Further studies are required to determine the exact nature of the interactions between ag85 proteins and the gelatin-binding site of human fibronectin.

## Disulfide bond isomerase—Rv2878c

Disulfide bonds are vital for protein folding and stability. In gram-negative bacteria, disulfide bonds are usually introduced in the periplasmic space by protein disulfide oxidoreductases.[70] However, since gram-positive bacteria do not have a periplasmic space, the disulfide oxidoreductases (Dsb) are secreted into the extracellular space surrounding the cell.[71,72] *M. tuberculosis* has two Dsb-like proteins (Rv2878c and Rv1677), both of which are secreted, and one "predicted" integral membrane protein (Rv2874) that is thought to complete the redox cycles of the two secreted Dsb-like proteins. Rv2878c has sequence homology to *E. coli* DsbE. *E. coli* DsbE has been shown to be a weak reductant which is involved in cytochrome c maturation in the periplasm.[73]

The crystal structure of *M. tuberculosis* DsbE reveals one domain that contains a thioredoxin fold,[74] Fig. 6. The structure shows the cysteines of the active site Cys-X-X-Cys in their reduced state[75] (Fig. 6). The precise function of *M. tuberculosis* DsbE is unknown. However, preliminary experimental evidence indicates that *M. tuberculosis* DsbE is an oxidant rather than a weak reductant, as is *E. coli* DsbE.[75] Therefore one could speculate that *M. tuberculosis* DsbE could potentially protect mycobacteria from oxidative damage by

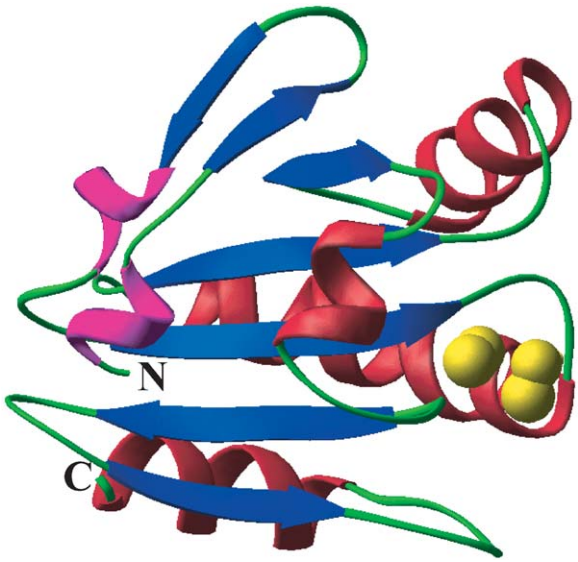



**Fig. 6** Three-dimensional structure of *M. tuberculosis* DsbE. Ribbon diagram of the monomer, $\alpha$-helices, $3_{10}\alpha$-helices and $\beta$-strands are shown in red, purple and blue, respectively. The active site sulfur and $\beta$-carbon atoms are shown in yellow. This figure was generated using RIBBONS.

macrophages and/or correct incorrectly formed disulfide bonds in secreted proteins.

### FolB—Rv3607c

Folate derivatives are essential cofactors in the biosynthesis of purines, pyrimidines and amino acids. Mammalian cells are able to utilize preformed folates after uptake by a carrier-mediated active transport system. Most microbes, such as *Mtb*, lack this system and must synthesise folates de novo from guanosine triphosphate.[76] We are studying 7,8-dihydroneopterin aldolase in the biosynthetic pathway of folic acid to explore the possibility of this protein as a drug target.

Dihydroneopterin aldolase (FolB) is an enzyme in the pathway of the biosynthesis of folic acid that catalyzes the conversion of 7,8-dihydroneopterin to 6-hydroxymethyl-7,8-dihydropterin. The structure is extremely similar to that of FolB from *Staphylococcus aureus*[77] and FolX from *E coli*.[78] The structure is an octamer, which consists of a dimer of tetramers that appears to have a pterin compound bound to its active site,[79] Fig. 7a. The monomers, which have an ellipsoidal shape, come together to form a 16-stranded anti-parallel $\beta$-barrel from four of the monomers, and is surrounded by a ring of $\alpha$-helices. The catalytic site is on the interface between two of the monomers. One of the more interesting features of this structure is that the electrostatic surface potential of *M. tuberculosis* FolB is more negative than that of *S. aureus* FolB, and the pore of the *S. aureus* FolB structure is more positively charged than that of *M. tuberculosis* FolB, Fig. 7b.

### Secreted protein—Rv1926c

Rv1926c is a major secreted protein of unknown function, specific to mycobacteria. It has been shown to stimulate humoral immune responses in guinea pigs infected with virulent *M. tuberculosis*.[80] T-cell epitope mapping showed that Rv1926c contains a highly immunodominant region at its N-terminus, although this may be the region of the signal peptide.[81] Using a Tn*552′pho*A in vitro transposition system, Braunstein et al. confirmed that Rv1926c is a secreted protein, and have shown it to be a cell-envelope-associated protein that may participate in virulence.[82] The structure of Rv1926c is a beta-sandwich fold consisting of two anti-parallel $\beta$-sheets, except for the N-terminal $\beta$-strand which forms a parallel $\beta$-sheet with the C-terminal $\beta$-strand,[83] Fig. 8.

The fold of Rv1926c is a member of the immunoglobulin superfamily.[84] Rv1926c has structural similarity to cell-surface binding proteins such as bovine arrestin, an intracellular inhibitory protein that terminates receptor activity,[85] mouse adaptin, a dimeric protein which binds to the trimer of the clathrin buds[86] and *Yersinia pseudotuberculosis* invasin, a bacterial protein which mediates entry into eukaryotic cells by binding to members of the $\beta_1$ integrin family.[87] Structural similarity implicates Rv1926c in possible host–bacterial interactions for endocytosis or phagocytosis. Genetic knock-out studies of the Rv1926c gene in *M. tuberculosis* are in progress to test for its role in infection.

## Chaperonin-10 structure and its relevance as antigen and in diseases (B. Taneja and S.C. Mande)

The chaperonins are important biomolecules that facilitate the folding of a variety of polypeptides in vivo in an ATP-dependent manner.[88] Initially discovered by their requirement for the correct assembly of the head proteins of phage lambda in mutant *E. coli*,[89] the function of chaperonins as molecular chaperones was first identified by their role in the assembly of oligomeric ribulose bisphosphate carboxylase in plant chloroplasts.[90]

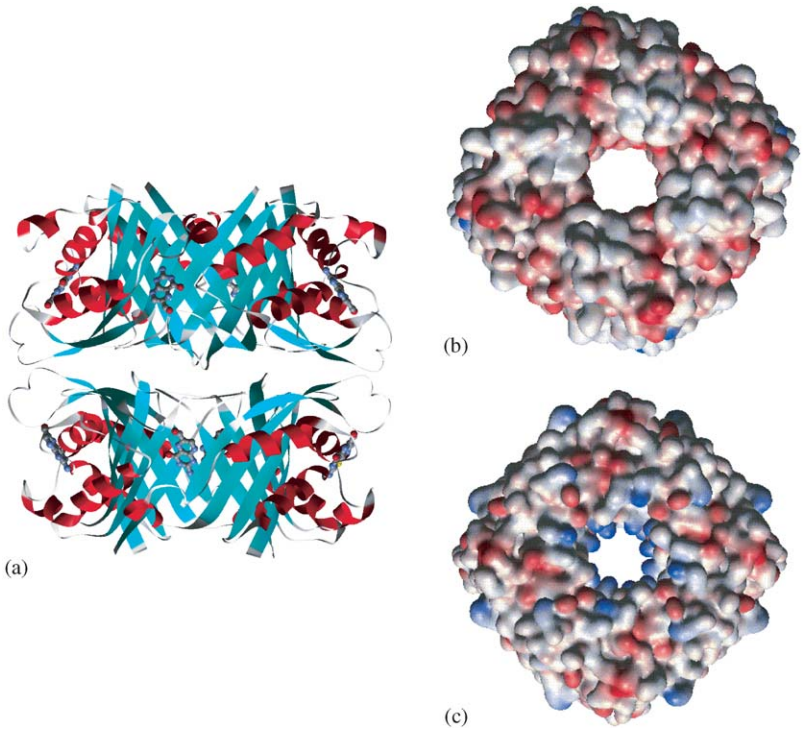



**Fig. 7** Three-dimensional structure of *M. tuberculosis* FolB. (a) The ribbon diagram of the structure of FolB is an octamer, which consists of a dimer of tetramers that appears to have a pterin compound bound to its active site, which is on the interface of two of the monomers. (b) and (c) The electrostatic surface potential of *M. tuberculosis* FolB (b) is more negatively charged than that of *S. aureus* FolB (c). The ribbons diagrams and electrostatic surface potentials were produced with WEBLAB VIEWER PRO 3.7.

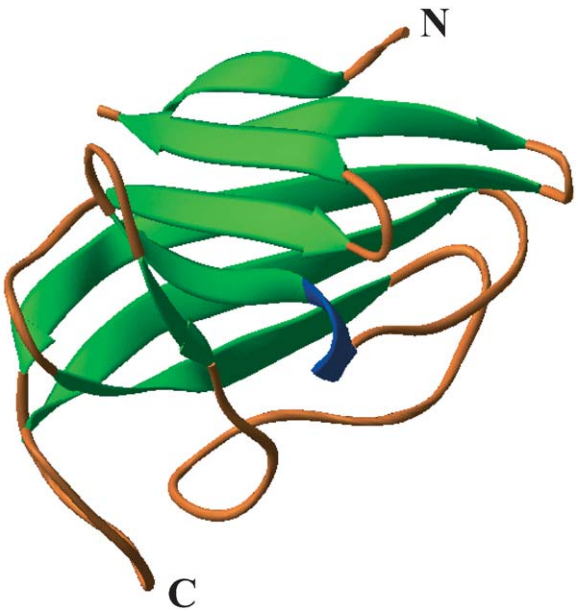



**Fig. 8** Three-dimensional structure of *M. tuberculosis* Rv1926c. The ribbon diagram of the structure of Rv1926c is similar to an immunoglobulin fold with the N- and C-termini $\beta$-strands forming a parallel $\beta$-sheet.

The chaperonins have subsequently been observed in various cellular compartments of both prokaryotes and eubacteria.

The chaperonins consist of two members, a larger 60 kDa subunit and a smaller 10 kDa subunit. Chaperonin-60 (cpn60) forms a large tetradecameric structure of identical subunits arranged as two heptameric rings stacked back to back. Each ring encloses a large central cavity where the substrate polypeptide is believed to bind. The function of cpn60 in the protein folding cycle is modulated by its cofactor, the homo-heptameric chaperonin-10 (cpn10), which acts as an allosteric modulator of chaperonin function. Binding of cpn10 and ATP results in large *en bloc* movements of cpn60 domains leading to a larger hydrophilic cavity in cpn60, conducive for folding of the substrate polypeptide.[91]

The crystal structure of *M. tuberculosis* cpn10 (Mtcpn10) is very similar to the *E. coli* GroES and *M. leprae* chaperonin-10 structures.[31,42,92] The structure of Mtcpn10 resembles a dome-like architecture with a small orifice at the top of 8–12 Å.[31,32] The overall structure of individual subunits consists of a $\beta$-barrel core conforming to the GroES-fold.[93,94] Cpn10 consists of two enigmatic regions, an approximate 10-residue region extending from the top of the dome and termed as the "dome loop" and a larger 17-residue stretch, extending from the bottom of the $\beta$-barrel core and termed as the "mobile loop".

The top of the dome loop is the closest approach to the seven-fold symmetry axis in the chaperonin-10 structures. In a large majority of cpn10 sequences, at least one of the residues at the tip of the dome is acidic in nature. The conglomeration of acidic residues and the resulting high negative potential at the tip of the dome has been shown to be responsible for the flexibility of the dome loop and the plasticity of the oligomeric structure in Mtcpn10.[95] A large increase in the diameter of the dome orifice of *E. coli* cpn10, as observed by small angle X-ray scattering experiments,[96] might also be due to this clustering of negative charges at the dome.

The crystal structure of Mtcpn10 shows divalent cations bound to at least one preferential binding site in all of the seven subunits of Mtcpn10.[32] Trp50 at the dome loop of Mtcpn10, seems to be involved in metal binding through its indole nitrogen along with Glu9 of the same subunit and Asp61 of the neighboring subunit (Fig. 9). Metal ions thus might bind to Trp and its neighboring residues resulting in decreased flexibility of the dome loop. The importance of divalent cations and their possible role in the stabilization of oligomeric states of cpn10 is also emphasized by the vast improvement in diffraction upon addition of divalent cations to the initial crystallization conditions of Mtcpn10.[31] Metal binding to cpn10 and the strict requirement of divalent cations for the cpn60:cpn10 complex formation also suggests interesting roles of metal ions as regulators of chaperonin complex formation.

Cpn10 or the 10 kDa antigen of both *M. tuberculosis* and *M. leprae* elicit a strong humoral and cell-mediated immune response upon infection. Specific immune responses to cpn10 include production of antibodies, T-cell proliferation and delayed type hypersensitivity.[97–99] Immunodominant epitopes from cpn10 of both *M. leprae* and *M. tuberculosis* have been identified.[100,101] These studies using truncated overlapping peptides of the entire cpn10 sequence have mapped the strongly immunodominant epitopes to the mobile loop.

The chaperonin-10 of *M. tuberculosis* has also been postulated to be one of the major factors responsible for bone resorption in Pott's disease.[102] These studies using synthetic overlapping peptides have identified the mobile loop region and the region comprising residues 65–70 to function as a single conformational unit to cause bone resorption. In the recent Mtcpn10 structure,[31] the mobile loop appears to be present as a single conformational unit with the residues 65–70. It is hence possible that these two regions of the secreted Mtcpn10 possess osteolytic activity as proposed earlier.[102]

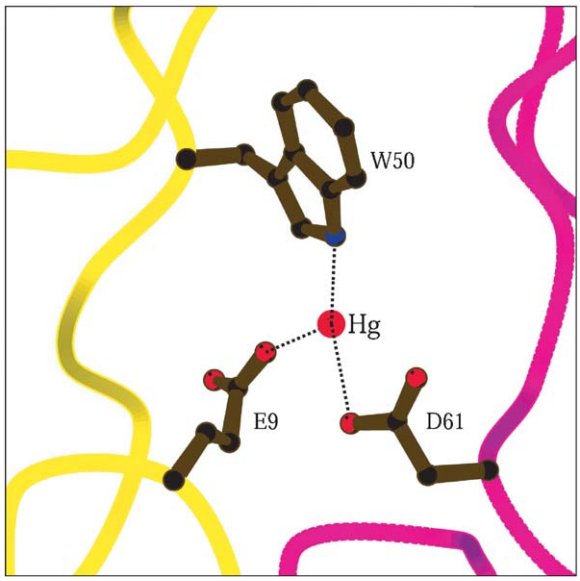



**Fig. 9** Metal ions bind to the dome loop of Mtcpn10: Trp50 at the dome loop of Mtcpn10 seems to be involved in binding of mercury along with Glu9 of the same subunit (yellow) and Asp61 of a neighboring subunit (purple).[30] Other metal ions thus might also bind to Trp and its neighboring residues resulting in decreased flexibility of the dome loop.

As much as 20% of the total protein content of culture filtrates of logarithmically growing *M. tuberculosis* has earlier been shown to consist of cpn10,[103] leading to the postulation of an alternative interesting possibility for the bone-resorbing activity of Mtcpn10. Specific chelation of calcium by the secreted cpn10 at bone joints could possibly disturb the equilibrium of calcium available for bone calcification and hence lead to bone deformities.

Inhibition of calcium-dependent signaling pathways is used by *M. tuberculosis* for its increased survival in the macrophages.[104] Hence chelation and subsequent depletion of calcium ions by Mtcpn10 could be a possible mechanism used by the pathogen for its increased intracellular survival. Alternatively the change in $Ca^{2+}$ levels might lead to $Ca^{2+}$ signaling pathways to manifest bone breakdown in Pott's disease.

The currently available structures of Mtcpn10[31,32] provide an attractive tool for understanding the intricate and complex functions of the 10 kDa antigen. Further biochemical work must be carried out to test the various possible activities of the chaperonin.

## Structural genomics of *M. tuberculosis* at the EMBL Hamburg Outstation (Ehmke Pohl, Victor Lamzin, Paul Tucker, Matthias Wilmanns, Christos ColovosWolfram Meyer-Klaucke)

The EMBL Hamburg Outstation joined the TB Structural Genomics Consortium in 1999. The Outstation is located at the German Synchrotron Facility DESY and operates seven beam lines for structural biology. Five of these beam lines are designed for protein crystallography and two of these are suitable for single/multiple anomalous diffraction (SAD/MAD) experiments.[105] The Outstation offers access to its beam line facilities and provides scientific support during the collection of diffraction data for the Consortium. In the near future, we will also offer the possibility of collecting data from frozen protein crystals sent to the EMBL Outstation by collaborating partners. We anticipate that the European members of the Consortium will primarily make use of the EMBL facilities. Additionally, researchers from the Laboratory have targeted a small number of *M. tuberculosis* genes for structure elucidation. These genes fall into three basic categories. (i) DNA-binding proteins that are believed to regulate essential metabolic functions and virulence factors in *M. tuberculosis*. One example is the ferric-uptake regulator (Fur). The functionally related iron-dependent regulator (IdeR) is implicated in the transcriptional control of genes involved in iron acquisition and storage, and genes for the survival of the pathogen in macrophages. The crystal structure of IdeR was solved even before the consortium started.[106] The two Fur orthologs are involved in the oxidative stress response and FurA has been shown to regulate the virulence of *M. tuberculosis* through catalase-peroxidase expression.[107] The *fur* genes have been cloned, expressed and purified by Celia Goulding at the University of California and crystallization experiments in collaboration with the EMBL Hamburg Outstation are underway.[108] (ii) Proteins from the histidine, lysine and leucine biosynthesis pathways. These proteins are of high general interest for the understanding of basic metabolic pathways in *Mtb*. In addition selected enzymes from these pathways constitute potential drug targets.[109] The crystal structures of a number of enzymes from the histidine biosynthesis pathway from hyperthermophilic bacteria have been elucidated[110] and will be used to assist in the structure determination of *M. tuberculosis* orthologs. (iii) Proteins of the various two-component systems. The prototype of this system consist of a sensor kinase and a response regulator. The two major domains of the sensor protein are an input domain that reacts to a specific stimulus, e.g. osmotic pressure or phosphate starvation, leading to autophosphorylation of the kinase domain. Subsequently, the kinase domain activates the DNA-binding domain via phosphorylation of the response element thus activating the DNA-binding domain. Members of this family have been implicated in the survival of the pathogen in the macrophage. So far, ten putative members of these systems have been cloned in the collaboration with the protein expression unit at the University of California, Los Angeles. Five of these proteins have been overexpressed and purified for crystallization experiments.[111]

In the future we will extend the target list using the proteomics techniques developed by the group headed by S.H.E. Kaufmann from the Max-Planck Institute for Infection Biology, Berlin.[112] In this approach, entire proteomes of different strains (e.g. virulent and avirulent) or bacterial cultures grown under different conditions are compared by two-dimensional electrophoresis (non-equilibrium pH gradient gel electrophoresis and SDS–PAGE). Distinct spots are then identified from the 2D-gel by peptide mass fingerprinting using MALDI-MS. The results are stored and updated in a dynamic data base on the World Wide Web (http://www.mpiib.berlin.mpg.de/2D-PAGE).[113] In collaboration with the Max-Planck Institute we plan to identify new targets for our structural studies. Knockout *M. tuberculosis* mutants will be generated to determine the biological role of the targeted proteins—emphasizing their role in virulence and drug resistance. This multidisciplinary strategy allows the full postgenomic analysis ranging from structure to function.

## Cytochrome P450 redox systems in *M. tuberculosis* (Andrew W. Munro, Kirsty J. McLean, Ker R. Marshall and David Leys)

Among the most notable features of the genome sequence of *M. tuberculosis* determined by Stewart Cole and co-workers[9] are the large proportion of the coding capacity given over to production of enzymes involved in lipid metabolism, and in particular the plethora of cytochrome P450 systems (P450s). The P450s are heme *b*-containing enzymes that bind and reduce molecular oxygen, leading to (usually) mono-oxygenation of substrate and production of a molecule of water. The large number

**Table 2** Cytochrome P450-encoding genes in *M. tuberculosis* H37RV.

| Gene | Genome locus (bp) | Predicted Mr (Da) | Comments |
|---|---|---|---|
| Rv0136 | 163366 | 49261 | Similar to other *M. tuberculosis* P450s |
| Rv0327c | 392699 | 50011 | Similar to *M. tuberculosis* Rv0568 |
| Rv0568 | 659450 | 50688 | Similar to other *M. tuberculosis* P450s |
| Rv0764c | 856685 | 50878 | Structurally characterized |
| Rv0766c | 858867 | 45422 | Similar to *S. erythraea* CYP107B1 |
| Rv0778 | 871431 | 45955 | Similar to *Streptomyces* nikkomycin synthesis P450s |
| Rv1256c | 1403389 | 44581 | Low similarity to known P450s |
| Rv1394c | 1569857 | 52230 | Similar to eukaryotic fatty acid hydroxylases |
| Rv1666c | 1892273 | 47865 | Similar to *Anabaena* P450s |
| Rv1777 | 2010656 | 47219 | Similar to *Streptomyces* narbomycin hydroxylase |
| Rv1785c | 2023450 | 43542 | Similar to *Rhizobium* P450s |
| Rv1880c | 2130544 | 48873 | Binds fatty acids |
| Rv2266 | 2540104 | 47825 | Similar to *Pseudomonas* linalool mono-oxygenase |
| Rv2268c | 2542810 | 53314 | Weak similarity to *S. erythraea* P450s |
| Rv2276 | 2547749 | 43257 | Structurally characterized |
| Rv3059 | 3419491 | 56228 | Putative sterol demethylase |
| Rv3121 | 3486508 | 43731 | Similar to *Streptomyces* mitomycin biosynthesis P450 |
| Rv3518c | 3954326 | 44399 | Similar to *Streptomyces* nikkomycin synthesis P450s |
| Rv3545c | 3984145 | 48433 | Similar to *M. tuberculosis* Rv2266 |
| Rv3685c | 4127297 | 52266 | Similar to other *M. tuberculosis* P450s |

All have predicted mass in the range 43–54 kDa, and all possess a signature amino acid sequence near the end of the protein indicating heme ligation by a conserved cysteinate: [FW]-[SGNH]-x-[GD]-x-[RKHPT]-x-C-[LIVMFAP]-[GAD].

of P450s in *M. tuberculosis* and the importance of lipid metabolism in the pathogen may be inextricably linked, since eukaryotic P450s typically have roles in lipid metabolism—including oxidation of fatty acids, eicosanoids, sterols and steroids.[114] It seems likely that the majority of the P450 systems in *M. tuberculosis* will have roles in lipid metabolism. These roles are likely to include synthesis and inter-conversions of fatty acid and mycolipids, polyketides and (possibly) sterols.

Cytochromes P450 are relatively rare in prokaryotes by comparison with eukaryotes. For instance, *E. coli* has no P450s. The *Bacillus subtilis* genome encodes seven P450s, a high number for a bacterium.[115] In contrast, mammals typically have more than 40 P450s, while the genome of *Arabidopsis thaliana* indicates more than 200 P450s. The genome sequence of *M. tuberculosis* indicates that there are likely to be 20 different cytochromes P450 produced by the bacterium (Table 2). At the time that the sequence was determined, this was far more than in any other bacterium. However, the *Streptomyces coelicolor* genome also predicts there to be more than 18 P450s, suggesting that the Actimomycetes may require large numbers of P450s for metabolic functions. The genome of *M. bovis* is highly similar to that of *M. tuberculosis* H37RV, although homologues of genes for two of the P450s from *M. tuberculosis* (encoded by Rv1256 and Rv3121) are absent due to deletion of relevant regions from the *M. bovis* genome. Interestingly, the genome of the leprosy pathogen *M. leprae* retains only a single P450.[17] However, in comparison with *M. tuberculosis*, numerous metabolic activities have been eliminated in *M. leprae*. Also, less than half the genome of *M. leprae* contains functional genes and pseudogenes (including 11 P450 pseudogenes) are abundant. The one functional P450 gene remaining in *M. leprae* is encoded by the ML2088 gene, the homologue of the Rv1880c-encoded P450 in Mtb. It may be the case that this P450 performs an essential catalytic function common to both species.

In order for catalytic function, P450s need to interact with flavoprotein or flavoprotein/ferredoxin redox partners which provide them with the NAD(P)H-derived electrons required for oxygen reduction and scission.[116] In eukaryotes, the redox partner is usually a single diflavin enzyme: NADPH-cytochrome P450 reductase. However, in the characterized bacterial P450 redox systems there are generally two redox partner enzymes involved: a FAD-containing NAD(P)H flavodoxin/ferredoxin reductase (FDR) and an iron-sulfur ferredoxin (or possibly an FMN-containing flavodoxin) (Fig. 10). The *M. tuberculosis* genome sequence encodes several predicted ferredoxins, but apparently no flavodoxins. Among these, the ferredoxin encoded

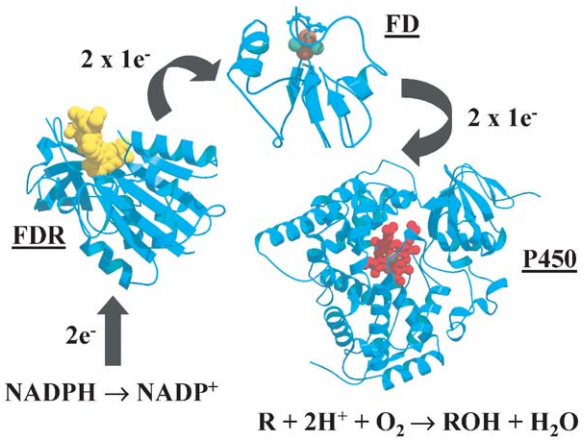



**Fig. 10** Scheme for the electron transfer pathway in *M. tuberculosis* P450 redox systems. Electrons for the P450 mono-oxygenation reaction are provided ultimately by NAD(P)H, which reduces an FAD-containing ferredoxin reductase (*FDR*) by hydride ion transfer. Electrons are shuttled one at a time to the iron sulfur cluster of a ferredoxin (*FD*), which (in turn) passes the electrons singly to the *P450*. The two-electron reduction of the P450 results in reductive scission of dioxygen and the hydroxylation of substrate (R) to product (ROH) and the formation of a molecule of water.

by the Rv0763c gene is located adjacent to a cytochrome P450-encoding gene (Rv0764c), suggesting that this may be a redox partner for the relevant P450s. The FprA gene encodes a FDR enzyme, while the FprB and FdxB genes are predicted to encode FDR/ferredoxin fusion enzymes, are possible single component P450 redox partners.

In advance of enzymatic characterization, it is difficult to predict the biochemical roles for several of the *M. tuberculosis* P450s. Frequently, the extent of amino acid sequence identity with characterized P450s provides strong clues to the nature of substrate preference for a novel P450. In fact, P450 sequences are organized into an extensive "gene superfamily", with P450s having $\geqslant$40% amino acid sequence identity classified as members of the same family and typically exhibiting selectivity for similar substrate classes.[117] However, the majority of the *M. tuberculosis* P450s show closest similarity to other (uncharacterized) P450s and/or have much less than 40% identity to prokaryotic P450s of known function. This may point to novel metabolic roles and unusual substrates for many of these *M. tuberculosis* P450s. Evidently, much work is required to establish the metabolic functions of these enzymes. Despite uncertainty as to the roles of many of the P450s, there are notable exceptions for which substrate selectivity appears more obvious. For instance, the product of the Rv0764c gene (CYP51) has highest similarity to eukaryotic sterol demethylase P450s, and catalyses 14$\alpha$-demethylation of lanosterol and the plant sterol obtusifoliol.[118] Sterol demethylase P450s are essential for synthesis of membrane sterols in eukaryotes, and the fungal CYP51 enzymes are targeted by azole drugs (e.g. fluconazole, clotrimazole). While definitive proof for a sterol biosynthetic pathway in the prokaryote is still lacking, the *M. tuberculosis* CYP51 certainly has capacity to demethylate eukaryotic sterols and presumably acts in vivo on substrates of similar structure. Also, the P450 product of the *M. tuberculosis* Rv1394c gene has strong similarity to mammalian family 4 fatty acid hydroxylases and is predicted to have specificity for long chain fatty acids. Genetic location may also provide clues to metabolic function of selected *M. tuberculosis* P450s, and in this respect the location of the Rv1880c gene in an operon including a potential mycolyl transferase gene (FpbB) may indicate that this P450 also has selectivity for fatty acids or mycolipids. Preliminary studies from our group with the purified P450 shows that it binds tightly to a range of long chain fatty acids.

While work on the *M. tuberculosis* P450 redox systems is still in its infancy, there have already been significant breakthroughs. The atomic structure of the CYP51 has been solved in the presence of the azole anti-fungal drug fluconazole, and also in another azole-bound form (4-phenylimidazole).[119] Recently, our group has also solved the structure of the P450 encoded by the Rv2276 gene (CYP121). This P450 shows greatest structural similarity to P450 eryF from *Saccharopolyspora erythraea*, involved in the synthesis of the polyketide antibiotic erythromycin. Both CYP51 and CYP121 bind very tightly to azole anti-fungal drugs, and these compounds also show good antibiotic properties in preliminary studies with *M. smegmatis* and *M. paratuberculosis*.[120] It appears that an anti-P450 strategy may be an effective and largely unexplored route to novel antibiotics for Mtb. In other work, our group has cloned and expressed several other *M. tuberculosis* P450s and redox partners, including the ferredoxin reductase FprA, which is likely to be the electron donor to ferredoxin partners for several of the *M. tuberculosis* P450s (see also section on FprA). The purified FprA exhibits NADPH-dependent reductase activity typical of bacterial and eukaryotic ferredoxin reductase enzymes supporting other P450 systems. In the course of the next few years, the roles of the individual P450 isoforms in *M. tuberculosis* biology will be elucidated, as will their relative importance

to the physiology of the pathogen and their potential as drug targets. With drug- and multi-drug-resistant strains of *M. tuberculosis* spreading across the globe, the desperate need for novel antibiotics is universally recognized. In this respect, the P450s are already validated anti-fungal drug targets, and with 20 different isoforms in *M. tuberculosis* it is virtually certain that a number of these serve indispensable metabolic functions. Thus, systematic evaluation of the biochemistry, enzymology and structure of the *M. tuberculosis* P450s could lay the foundations required for a new and effective strategy to counter the pathogen.

## A structural genomics study on *M. tuberculosis* at Seoul National University (Jin Kuk Yang, Hye-Jin Yoon, Byung Il Lee, Myong Gyong Lee, Jae Eun Kwak, Byung Woo Han, Jae Young Lee, Seung-Hun Baek, Min S. Park, Geoffrey S. Waldo and Se Won Suh)

As part of the structural genomics project on *M. tuberculosis*, we have targeted and studied 45 genes. Most of our target genes are involved in the synthesis of amino acids, fatty acids, cofactors, and prosthetic groups. They have been cloned into pET expression vectors (Novagen), for expressing the target proteins in either intact form or N- or C-terminal histidine-tagged form. When we expressed the proteins in *E. coli* cells (BL21(DE3), C41(DE3), or B834(DE3)), 19 were not expressed, 19 were expressed in an insoluble form as inclusion bodies, and seven were expressed and were highly soluble. Using a green fluorescent protein-based directed evolution method,[121] we made a soluble mutant of the Rv2002 gene product, which was initially expressed in an insoluble form within inclusion bodies. We have so far successfully purified six of these *M. tuberculosis* proteins: Rv0254c (cobU), Rv2002 (fabG3) I6T/V47M/T69 K mutant, Rv2043c (pncA), Rv2534c (efp), Rv2538c (aroB), and Rv3420c (rimI). The Rv2002 soluble mutant was crystallized[122] and its crystal structure was determined at 1.8 Å resolution. It reveals that Rv2002 is likely to function as NADH-dependent short-chain dehydrogenase/reductase. Further experiments are necessary to establish its biological role. Rv2538c (aroB), encoding 3-dehydroquinate synthase of the shikimate pathway, has been crystallized but the crystals are too small for X-ray diffraction analysis. Further optimization of its crystallization condition is in progress. Rv2043c (pncA) encodes pyrazinamidase/nicotinamidase and its mutation is known to cause resistance to antituberculosis drug pyrazinamide.[123] It was over-expressed in either intact form or with a C-terminal hexa-histidine tag. The purified protein with the C-terminal tag was monomeric as determined by dynamic light scattering measurements, showing just 21% polydispersity. Metal analysis by Inductively Coupled Plasma mass spectrometry revealed that each mole of protein contained 0.88 mole of Fe and 0.25 mole of Ni. Despite extensive crystallization trials, we have not succeeded in crystallizing it so far.

## Characterization of isoniazid-induced proteins in *M. tuberculosis* (Moyra M. Komen, Vickery L. Arcus, Edward N. Baker and J. Shaun Lott)

Isoniazid (INH) is thought to exert its bactericidal action on bacteria of the *M. tuberculosis* complex by inhibiting the biosynthesis of cell wall mycolic acids, thereby rendering the cell susceptible to reactive oxygen radicals and other environmental factors.[124] A recent DNA micro-array gene expression study on *M. tuberculosis* identified a cohort of genes that are upregulated in the bacterium immediately after exposure to INH.[125] A cluster of genes thought to be involved in mycolate biosynthesis and fatty acid metabolism, and one gene known to be involved in free radical detoxification (the alkyl hydroperoxide reductase *ahpC*) were among those identified as being upregulated. These findings therefore agree with the current knowledge on the mode of action of INH. A further set of genes annotated as 'of unknown function' in the genome sequence were also found to be upregulated. Our hypothesis is that some of the proteins encoded by this upregulated cohort of genes may be involved in a protective response by the bacterium to the cytotoxic consequences of INH activity.

The INH-induced cohort of genes has been screened using a variety of bioinformatics techniques in order to select those amenable for structural analysis using X-ray or NMR techniques. The preliminary functional and biophysical characterization of a few of these proteins is reviewed here.

### Rv1592c

This open reading frame was originally annotated in the genome as a ''conserved hypothetical'', but the

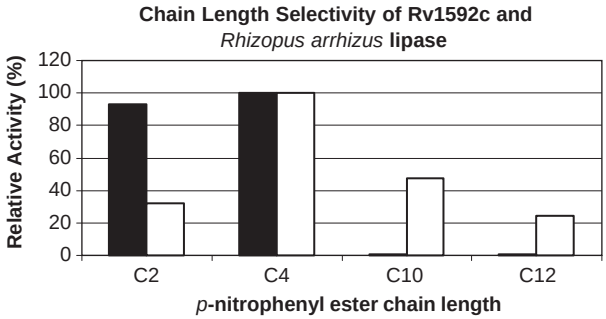



**Fig. 11** Chain length selectivity of Rv1592c (black bars) compared to the *Rhizopus arrhizus* lipase (Sigma L4384, white bars). Selectivity assays were performed on a range of *p*-nitrophenol ester substrates essentially as described by van Kampen et al.[129] Activities were normalized to make the activity on *p*-nitrophenol butyrate equal to 100%.

cloning of a family of lipases from *Candida albicans* has indicated that Rv1592c may also encode a lipase. Hube and co-workers[126] identified a ten-member gene family of secreted lipases (*LIP1-10*) in *C. albicans*, that encode proteins of ~50 kDa sharing between 33% and 80% amino acid sequence identity. When compared to other known fungal lipases, the *Candida* enzymes form a distinct clade, and are more similar in sequence to Rv1592c than they are to other fungal enzymes: Rv1592c and the *Candida* genes are ~30% identical in amino acid sequence. Rv1592c contains conserved active site residues, indicating that it may have lipase activity. The *LIP* gene family appears to be restricted to pathogenic species of *Candida*, and may play a role in persistence and virulence in those organisms. *C. albicans* shares with *M tuberculosis* the ability to survive in the macrophage phageosome, and it is not unreasonable to suppose that the two organisms may employ similar survival strategies in this context. It may be then that lipases such as Rv1592c play a role in mycobacterial persistence or virulence.[127] However, a key difference between Rv1592c and the *LIP* family lipases is that the former does not have a recognizable secretion signal.

Rv1592c has been expressed as an *N*-terminally His-tagged protein in the cytosol of *E. coli*. Under most conditions tested, the protein is insoluble. However, growing the expression cells at 18°C and lysing them in a pH 9 buffer yields some soluble protein. The solubility of the protein is markedly and selectively improved by the addition of specific detergents to the lysis buffer: either *n*-dodecyl-*n*-*n*-dimethylamino-3-propanesulfonate (a zwittergent) or *n*-dodecyl-$\beta$-D-maltoside (a non-ionic detergent) yield almost totally soluble protein. The protein can then be substantially purified in a single step using immobilized metal ion affinity chromatography.

We have assayed the lipase activity of the purified enzyme on a number of lipid substrates. Preliminary qualitative assays of lipolytic activity, monitored using rhodamine B in olive oil agar,[128] gave negative results. However, Rv1592c exhibited an esterase activity, measured spectrophotometrically using *p*-nitrophenol esters,[129] with a marked specificity for short ($C_4$–$C_4$) alkyl chain lengths (Fig. 11). Quantitative assays of the activity of the recombinant enzyme on tritiated triolein in an emulsified system[130] confirmed the inability of the enzyme to hydrolyse long-chain triglycerides.

The physiological function of Rv1592c remains uncertain. Its induction by INH and its lack of a signal sequence lead us to formulate two hypotheses regarding its likely role. Assuming that it is not cryptically secreted, the simplest explanation of its induction by INH treatment is that it is being used by the bacterium to hydrolyse the cytosolic pool of saturated fatty acids that accumulates due to the blockade of mycolic acid biosynthesis.[131] However, our preliminary substrate specificity studies make this seem unlikely, as the accumulating pool of fatty acid intermediates are long chain molecules ($C_{24}$–$C_{26}$). An alternative explanation is that Rv1592c is in some way involved in the removal of toxic lipid hydroperoxides from the cell, perhaps acting in concert with hydroperoxide reductases such as *ahpC*. Circumstantial evidence to support this notion comes from the observation that Rv1592c and *ahpC* are also upregulated by a similar amount (approximately threefold) when *M. tuberculosis* is shifted from ambient to reduced oxygen tension,[132] implying that both are in some way involved in mechanisms involved in the cellular adaptation to a shift in the redox environment. The elucidation of the precise function of this gene in the cell awaits the identification of its preferred physiological substrate(s). To this end, further specificity assays are being performed on the

purified enzyme and it is being subjected to crystallization trials, as the three-dimensional structure of the protein will likely hold useful clues to its cellular function.

## Rv0340-Rv0343

The open reading frames Rv0341 and Rv0342 were upregulated by INH in a delayed fashion in comparison to the others identified.[125] This delayed induction is also seen for *ahpC*, inferring that Rv0341 and Rv0342 may be involved in responding to the secondary, toxic results of INH activity. Another study looking at differential mRNA expression in *M. tuberculosis* treated with INH[133] also identified Rv0342 and Rv0343 as being strongly upregulated by INH and to a lesser extent by ethambutol (EMB), another antibiotic that inhibits cell wall biosynthesis via a distinct mechanism of action. Alland et al.[133] have termed Rv0341 as *iniB*, Rv0342 as *iniA* and Rv0343 as *iniC* respectively, and have subsequently investigated the promoter that appears to drive the operonic expression of these three genes.[134] The relatively slow induction of the *iniBAC* promoter compared to other INH-responsive genes was confirmed by this study: *iniBAC* is measurably induced after 4 h, and reaches a maximum between 24 and 48 h, which is the same timeframe that INH requires to cause cell death. The *iniBAC* operator was also found to respond to a range of antibiotics that act by inhibition of cell wall biosynthesis. Together, these observations imply that this gene cluster may be of key importance in cell viability when membrane integrity is being compromised by the action of antibiotics, and that they may represent good targets for the design of new antibiotics to act in concert with existing cell wall biosynthesis inhibitors such as INH and EMB.

None of the ORFs in this cluster show informative amino acid sequence similarity to any other proteins, and their biochemical functions remain unknown. Rv0342 (*iniA*) contains a phosphopantetheine attachment motif, leading to the suggestion that it may be an acyl carrier protein (ACP).[133] Although this makes biological sense, given that another ACP *acpM* (Rv2244) is both upregulated by and binds to INH, other consensus sequence motifs characteristic of ACPs are not found in the *iniA* sequence, which implies that *iniA* is either very dissimilar in sequence to known examples or does not encode an ACP. Despite their apparent lack of homology to other genes, the four genes in this cluster do show significant sequence similarity to each other. BLAST searches of the *M. tuberculosis* genone indicate that Rv0340 and Rv0341 form a paralogous pair of genes, as do Rv0342 and Rv0343. The two pairs of genes show clear conservation in particular segments of their sequences (Fig. 12), which suggests that they may share conserved domains which confer similar function, or which may be subunits of a larger macromolecular complex.

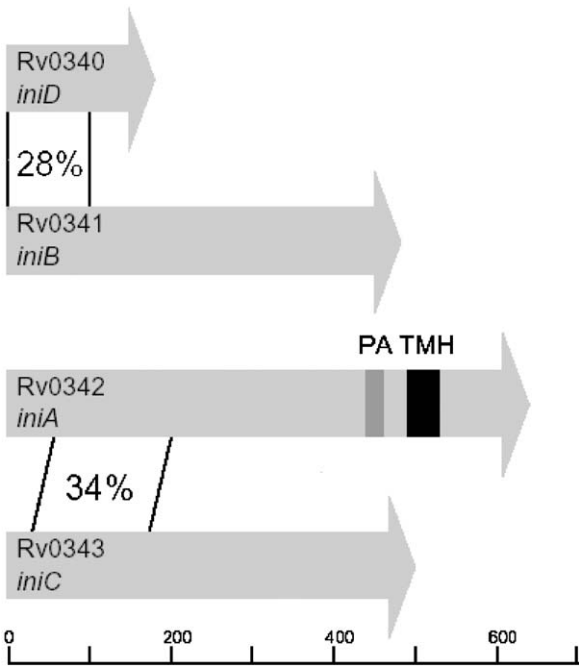



**Fig. 12** The paralogous relationship between ORFs in the *ini* operon. Percentage figures indicate the percentage identity of the sequences at the amino acid level in the conserved segments. The predicted trans-membrane helix (TMH) and phosphopantetheine attachment site (PA) in *iniA* are marked as colored bars. The scale indicates the length of the protein sequences, marked in hundreds of amino acids.

Although Rv0340 (which might be termed "*iniD*" for consistency with the other members of this gene cluster) does not appear to be induced along with the *iniBAC* operon, it has been chosen as a target for structural studies due to its sequence similarity to *iniB* and its relative tractability. Rv0340 has been expressed as an *N*-terminally His-tagged protein in *E. coli* and purified to homogeneity using a combination of immobilized metal affinity and size exclusion chromatography. It is an extremely soluble protein, and can be concentrated to higher than 40 mg ml$^{-1}$. However, attempts to crystallize it have failed, producing phase separations or oils over a wide and varied range of precipitant conditions, and NMR spectroscopy has confirmed that the protein is mainly unstructured in solution. This lack of tertiary

structure is perhaps not too surprising given the low complexity of the amino acid sequence of Rv0340 (15% alanine, 10% proline, 10% valine and 9% glycine). The conservation of sequence between Rv0340 and *iniB* in their first ~90 amino acids lead us to believe that this segment may constitute a discrete folded domain. We have confirmed the presence of the domain using limited tryptic digestion, mass spectrometry and amino acid sequencing. The isolated domain can be expressed and purified in a similar fashion to the full length protein, and NMR spectroscopy of the domain indicates that it does form a discretely folded structure in solution. We are currently working to solve the structure of this domain by high-field heteronuclear NMR spectroscopy, and are extending our successful domain definition strategy to the other members of this gene cluster.

Genomic techniques have enabled the identification of genes used by *M. tuberculosis* in its protective response to the toxic consequences of INH treatment. However, this assignment of physiological function gives no clue as to the biochemical mechanisms involved within the cell. We are using a combination of biochemical and biophysical methods to establish these mechanisms. When no useful inference about gene function can be drawn from sequence alone, protein structural analysis is a valuable, and perhaps the only viable, route to understanding the cytosolic functions of these protective gene products, and ultimately to controlling their activity with new therapeutic compounds.

## Acknowledgements

Work from the laboratory of Se Won Suh was supported by the Korea Ministry of Science and Technology (NRL-2001, grant no. M1-0104000132-01J000006210). Work from the laboratory of University of Pavia was supported by the Italian Ministry of University and Scientific and Technological Research (MURST, grant COFIN-1998, COFIN-2002) and by the European Union research project Quality of Life and Management of Living Resources (Contract No. QLK2-2000-01761). Work from many of the laboratories was supported by the NIH. The UCLA authors wish to acknowledge Mari Gingery, Duilio Cascio, Michael Sawaya and Cameron Mura for useful suggestions and discussion. The UCLA work has been supported by grants from the Department of Energy and the National Institutes of Health. The coordinates for Rv1886c, Rv2220, Rv1926c and Rv2878c have been deposited in the Protein Data Bank with the accession numbers 1FOP, 1HTO, 1LM1 and 1LU4, respectively. The SGHMS authors wish to thank Jon Cooper, Darren Thompson, Gerard Kleywegt and Peter Tormay for their help and advice, the Raymond Burton Medical Trust for a grant to A.R.M.C. and the Wellcome Trust for a grant to S.P.W. We would like to thank David Christie for originally suggesting this area of study at Auckland, Kerry Loomes for help and advice with lipase assays, the Auckland Medical Research Foundation for a project grant to V.A. and J.S.L, and the New Economy Research Fund and the Marsden Fund for project grants to E.N.B.

The coordinates for Rv3418c have been deposited in the Protein Data Bank with the accession number 1P3H.

**Note added in proof**

The chaperonin 10 work by Roberts et al has recently been published (Roberts MM et al., *J Bacteriol* 2003;**185**:4172–85).

## References

1. Burley SK. An overview of structural genomics. *Nat Struct Biol* 2000;**7**:932–4.
2. Brenner SE, Levitt M. Expectations from structural genomics. *Protein Sci* 2000;**9**:197–200.
3. Terwilliger TC, Waldo G, Peat TS, Newman JM, Chu K, Berendzen J. A protein structure initiative. *Protein Sci* 1998;**7**:1851–6.
4. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, Chan WT, Swenson D, Sacchettini JC, Jacobs Jr WR, Russell DG. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 2000;**406**:735–8.
5. Honer Zu Bentrup K, Miczak A, Swenson DL, Russell DG. Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J Bacteriol* 1999;**181**:7161–7.
6. Graham JE, Clark-Curtiss JE. Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences SCOTS. *Proc Natl Acad Sci USA* 1999;**96**:11554–9.
7. Sharma V, Sharma S, Hoener zu Bentrup K, McKinney JD, Russell DG, Jacobs Jr WR, Sacchettini JC. Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nat Struct Biol* 2000;**7**:663–8.
8. Huang C, Sacchettini J., in preparation, 2002.
9. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;**393**:537–44.
10. Diehl P, McFadden BA. Site-directed mutagenesis of lysine 193 in *Escherichia coli* isocitrate lyase by use of unique restriction enzyme site elimination. *J Bacteriol* 1993;**175**:2263–70.
11. Rehmann A, McFadden BA. Cysteine 195 has a critical functional role in catalysis by isocitrate lyase from *Escherichia coli*. *Curr Microbiol* 1997;**35**:267–369.

12. Rehman A, McFadden BA. Lysine 194 is functional in isocitrate lyase from *Escherichia coli*. *Curr Microbiol* 1997;**35**:14–7.
13. Rehman A, McFadden BA. Serine 319 and 321 are functional in isocitrate lyase from *Escherichia coli*. *Curr Microbiol* 1997;**34**:205–11.
14. Luttik MA, Kotter P, Salomons FA, van der Klei IJ, van Dijken JP, Pronk JT. The *Saccharomyces cerevisiae* ICL2 gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism. *J Bacteriol* 2000; **182**:7007–13.
15. Brock M, Darley D, Textor S, Buckel W. 2-Methylisocitrate lyases from the bacterium *Escherichia coli* and the filamentous fungus *Aspergillus nidulans*: characterization and comparison of both enzymes. *Eur J Biochem* 2001; **268**:3577–86.
16. Horswill AR, Escalante-Semerena JC. Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the prp locus at 8.5 centisomes, prpr encodes a member of the sigma-54 family of activators, and the prpbcde genes constitute an operon. *J Bacteriol* 1997;**179**:928–40.
17. Cole ST, Eiglmeier K, Parkhill J, et al. Massive gene decay in the leprosy bacillus. *Nature* 2001;**409**:1007–111.
18. Britton K, Langridge S, Baker PJ, Weeradechapon K, Sedelnikova SE, De Lucas JR, Rice DW, Turner G. The crystal structure and active site location of isocitrate lyase from the fungus *Aspergillus nidulans*. *Struct Fold Des* 2000;**8**:349–62.
19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;**22**:4673–80.
20. Magni G, Amici A, Emanuelli M, Raffaelli N, Ruggieri S. Enzymology of $NAD^+$ biosynthesis. *Adv Enzymol Relat Areas Mol Biol* 1999;**73**:135–82.
21. Rizzi M, Nessi C, Mattevi A, Coda A, Bolognesi M, Galizzi A. Crystal structure of NH3 dependent NAD synthetase from *Bacillus subtilis*. *EMBO J* 1996;**15**:5125–33.
22. Cantoni R, Branzoni M, Labò M, Rizzi M, Riccardi G. The *MTCY428.08* gene of *Mycobacterium tuberculosis* codes for $NAD^+$ synthetase. *J Bacteriol* 1998;**180**:3218–21.
23. Bellinzoni M, De Rossi E, Branzoni M, Milano A, Peverali F A, Rizzi M, Riccardi G. Heterologous expression, purification, and enzymatic activity of *Mycobacterium tuberculosis* $NAD^+$ synthetase. *Protein Expres Purif* 2002;**25**(3):547–57.
24. Bossi RT, Aliverti A, Raimondi D, Fischer F, Zanetti G, Ferrari D, Tahallah N, Maier CS, Heck AJR, Rizzi M, Mattevi A. A covalent modification of $NADP^+$ revealed by the atomic resolution structure of FprA, a *Mycobacterium tuberculosis* oxidoreductase. *Biochemistry* 2002;**41**(28): 8807–18.
25. Litwin CM, Calderwood SB. Role of iron in regulation of virulence genes. *Clin Microbiol Rev* 1993;**16**:137–49.
26. Zahrt TC, Song J, Siple J, Deretic V. Mycobacterial FurA is a negative regulator of catalase-peroxidase gene *katG*. *Mol Microbiol* 2001;**39**:1174–85.
27. Milano A, Forti F, Sala C, Riccardi G, Ghisotti D. Transcriptional regulation of *furA* and *katG* upon oxidative stress in *Mycobacterium smegmatis*. *J Bacteriol* 2001;**183**: 6801–6.
28. Rye HS, Roseman AM, Chen S, Furtak K, Fenton WA, Saibil HR, Horwich AL. GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* 1999;**97**:325–38.
29. Kong TH, Coates ARM, Butcher PD, Hickman CJ, Shinnick TM. *Mycobacterium tuberculosis* expresses two cpn60 homologs. *Proc Natl Acad Sci USA* 1993;**90**: 2608–12.
30. Roberts MM, Coker AR, Fossati G, Mascagni P, Coates ARM, Wood SP. Crystallization, X-ray diffraction and preliminary structure analysis of *Mycobacterium tuberculosis* chaperonin 10. *Acta Crystallogr D* 1999;**55**:910–4.
31. Taneja B, Mande SC. Three-dimensional structure of *Mycobacterium tuberculosis* chaperonin-10 reveals a partially stable conformation of its mobile loop. *Curr Sci* 2001; **81**:87–91.
32. Taneja B, Mande SC. Structure of *Mycobacterium tuberculosis* chaperonin 10 at 3.5 Å resolution. *Acta Crystallogr D* 2002;**58**:260–6.
33. Adams PD, Pannu NS, Read RJ, Brűnger AT. Cross-validated maximum likelihood enhances simulated annealing refinement. *Proc Natl Acad Sci USA* 1997;**94**:5018–23.
34. Rice LM, Brünger AT. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Struct Funct Genet* 1994;**19**:277–90.
35. Brünger AT. Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J Mol Biol* 1988;**203**: 803–16.
36. Brünger AT, Adams PD, Clore GM, et al. Crystallography & NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr D* 1998;**54**: 905–21.
37. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;**18**:2714–23.
38. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* 1994;**50**:760–763.
39. Kleywegt GJ, Jones TA. xdlMAPMAN and xdlDATAMAN, programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr D* 1996;**52**:826–8.
40. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 1971; **55**:379–400.
41. Hunt JF, Weaver AJ, Landry SJ, Gierasch L, Deisenhofer J. The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature* 1996;**379**:37–45.
42. Mande SC, Mehra V, Bloom BR, Hol WG. Structure of the heat shock protein chaperonin-10 of *Mycobacterium leprae*. *Science* 1996;**271**:203–7.
43. Hunt JF, van der Vies SM, Henry L, Deisenhofer J. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. *Cell* 1997;**90**:361–71.
44. Landry SJ, Zeilstra-Ryalls J, Fayet O, Georgopoulos C, Gierasch LM. Characterization of a functionally important mobile domain of GroES. *Nature* 1993;**364**:255–8.
45. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 1997;**388**:741–9.
46. Landry SJ, Taher A, Georgopoulos C, van der Vies SM. Interplay of structure and disorder in cochaperonin mobile loops. *Cell* 1996;**90**:361–71.
47. Landry SJ, Steede NK, Maskos K. Temperature dependence of backbone dynamics in loops of human mitochondrial heat shock protein 10. *Biochemistry* 1997; **36**:10975–86.

48. Laskowski RA, McArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the quality of protein structures. *J Appl Crystallogr* 1993;**26**:283–91.
49. Vaguine AA, Richelle J, Wodak SJ. *SFCHECK*: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D* 1999;**55**:191–205.
50. Luzzatti PV. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr* 1952;**5**:802–10.
51. Matthews BW. Solvent content of protein crystals. *J Mol Biol* 1968;**33**:491–7.
52. World Health Organization. The World Health report 1998: life in the 21st century—a vision for all. Geneva, Switzerland: The World Health Organization; 1998.
53. Guilhot C, Jackson M, Gicquel B, Mycobacteria: molecular biology and virulence. Blackwell Science Ltd, Oxford; 1999: p. 17–37.
54. http://www.doe-mbi.ucla.edu/TB/
55. Goulding CW, Perry LJ, Anderson DH et al. Progress at UCLA structural genomics of *Mycobacterium tuberculosis*: A preliminary report of progress at UCLA, *Biophysical Chem* 2003: in press.
56. Harth G, Horwitz MA. An inhibitor of exported *Mycobacterium tuberculosis* glutamine synthetase selectively blocks the growth of pathogenic mycobacteria in axenic culture and in human monocytes: extracellular proteins as potential novel drug targets. *J Exp Med* 1999;**189**: 1425–36.
57. Tullius MV, Harth G, Horwitz MA. High extracellular levels of *Mycobacterium tuberculosis* glutamine synthetase and superoxide dismutase in actively growing cultures are due to high expression and extracellular stability rather than to a protein-specific export mechanism. *Infect Immun* 2001; **69**:6348–63.
58. Harth G, Clemens DL, Horwitz MA. Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. *Proc Natl Acad Sci USA* 1994;**91**:9342–6.
59. Hirschfield GR, McNeil M, Brennan PJ. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. *J Bacteriol* 1990;**172**:1005–13.
60. Gill HS, Pfluegl GM, Eisenberg D. Preliminary crystallographic studies on glutamine synthetase from *Mycobacterium tuberculosis*. *Acta Crystallogr D Biol Crystallogr* 1999;**55**:865–8.
61. Gill HS, Pfluegl GMU, Eisenberg D. Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in enzymatic mechanism. *Biochemistry* 2002;**41**(31): 9863–72.
62. Gill HS, Eisenberg D. The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry* 2001;**40**:1903–12.
63. Naito M, Ohara N, Matsumoto S, Yamada T. The novel fibronectin-binding motif and key residues of mycobacteria. *J Biol Chem* 1998;**273**:2905–9.
64. Nardini M, Ridder IS, Rozeboom HJ, et al. The X-ray structure of epoxide hydrolase from *Agrobacterium radiobacter* AD1. An enzyme to detoxify harmful epoxides. *J Biol Chem* 1999;**274**:14579–86.
65. Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ, Besra GS. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* 1997;**276**: 1420–2.
66. Horwitz MA, Lee BW, Dillon BJ, Harth G. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1995;**92**:1530–4.
67. Ronning DR, Klabunde T, Besra GS, Vissa VD, Belisle JT, Sacchettini JC. Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nature Struct Biol* 2000;**7**:141–6.
68. Anderson DH, Harth G, Horwitz MA, Eisenberg D. An interfacial mechanism and a class of inhibitors inferred from two crystal structures of the *Mycobacterium tuberculosis* 30 kDa major secretory protein (Antigen 85B), a mycolyl transferase. *J Mol Biol* 2001;**307**:671–81.
69. Naito M, Fukuda T, Sekiguchi K, Yamada T. The domains of human fibronectin mediating the binding of alpha antigen, the most immunopotent antigen of mycobacteria that induces protective immunity against mycobacterial infection. *Biochem J* 2000;**347**:725–31.
70. Debarbieux L, Beckwith J. Electron avenue: pathways of disulfide bond formation and isomerization. *Cell* 1999;**99**:117–9.
71. Wiker HG, Michell SL, Hewinson RG, Spierings E, Nagai S, Harboe M. Cloning, expression and significance of MPT53 for identification of secreted proteins of *Mycobacterium tuberculosis*. *Microbial Pathog* 1999;**26**:207–19.
72. Gomez M, Johnson S, Gennaro ML. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect Immun* 2000;**68**:2323–7.
73. Li Q, Hu H, Xu G. Biochemical characterization of the thioredoxin domain of *Escherichia coli* DsbE protein reveals a weak reductant. *Biochem Biophys Res Commun* 2001; **283**:849–53.
74. Martin JL. Thioredoxin—a fold for all reasons. *Structure* 1995;**3**:245–50.
75. Goulding CW, Apostol MI, Parseghian A, Gennaro M, Eisenberg D. 1.1 Å resolution crystal structure of a secreted *Mycobacterium tuberculosis* disulfide oxidoreductase homologous to *E. coli* DsbE: implications for function, 2003, in preparation.
76. Achari A, Somers DO, Champness JN, Bryant PK, Rosemond J, Stammers DK. Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nature Struct Biol* 1997;**4**:490–7.
77. Hennig M, D'Arcy A, Hampele IC, Page MG, Oefner C, Dale GE. Crystal structure and reaction mechanism of 7,8-dihydroneopterin aldolase from *Staphylococcus aureus*. *Nature Struct Biol* 1998;**5**:357–62.
78. Ploom T, Haussmann C, Hof P, et al. Crystal structure of 7,8-dihydroneopterin triphosphate epimerase. *Struct Fold Des* 1999;**7**:509–16.
79. Goulding CW, Apostol MI, Parseghian A, Sawaya M, Eisenberg D. Crystal structure of FolB—implications for drug design, 2003, in preparation.
80. Manca C, Lyashchenko K, Wiker HG, Usai D, Colangeli R, Gennaro ML. Molecular cloning, purification, and serological characterization of MPT63, a novel antigen secreted by *Mycobacterium tuberculosis*. *Infect Immun* 1997;**65**:16–23.
81. Lee BY, Horwitz MA. T-cell epitope mapping of the three most abundant extracellular proteins of *Mycobacterium tuberculosis* in outbred guinea pigs. *Infect Immun* 1999; **67**:2665–70.
82. Braunstein M, Griffin TI, Kriakov JI, Friedman ST, Grindley ND, Jacobs Jr WR. Identification of genes encoding exported *Mycobacterium tuberculosis* proteins using a Tn552′phoA in vitro transposition system. *J Bacteriol* 2002;**182**:2732–40.

83. Goulding CW, Parseghian A, Sawaya MR, Cascio D, Apostol MI, Gennaro ML, Eisenberg D. Crystal structure of a major secreted protein of *Mycobacterium tuberculosis*—MPT63 at 1.5-A resolution. *Protein Sci* 2002;**11**(12):2887–93.
84. Halaby DM, Mornon JP. The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J Mol Evol* 1998;**46**:389–400.
85. Hirsch JA, Schubert C, Gurevich VV, Sigler PB. The 2.8 A crystal structure of visual arrestin: a model for arrestin's regulation. *Cell* 1999;**97**:257–69.
86. Owen DJ, Vallis Y, Pearse BM, McMahon HT, Evans PR. The structure and function of the beta 2-adaptin appendage domain. *EMBO J* 2000;**19**:4216–27.
87. Hamburger ZA, Brown MS, Isberg RR, Bjorkman PJ. Crystal structure of invasin: a bacterial integrin-binding protein. *Science* 1999;**286**:291–5.
88. Hartl FU. Molecular chaperones in cellular protein folding. *Nature* 1996;**381**:571–80.
89. Sternberg N. Properties of a mutant of *Escherichia coli* defective in bacteriophage head formation (*groE*). I. Initial characterization. *J Mol Biol* 1973;**76**:1–23.
90. Hemmingsen SM, Woolford C, van der Vies SM, et al. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 1988;**333**:330–4.
91. Xu Z, Horwich AL, Sigler PB. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 1997;**388**:741–50.
92. Hunt JF, Weaver AJ, Landry SJ, Gierasch L, Deisenhofer J. The crystal structure of GroES co-chaperonin at 2.8 Å resolution. *Nature* 1996;**379**:37–45.
93. Murzin AG. Structural classification of proteins: new superfamilies. *Curr Opin Struct Biol* 1996;**6**:386–94.
94. Taneja B, Mande SC. Conserved structural features and sequence patterns in the GroES fold family. *Protein Eng* 1999;**12**:815–8.
95. Taneja B, Mande SC. Metal ions modulate the plastic nature of *Mycobacterium tuberculosis* chaperonin-10. *Protein Eng* 2001;**14**:391–5.
96. Timchenko AA, Melnik BS, Kihara H, Kimura K, Semisotnov GV. GroES co-chaperonin small angle X-ray scattering study shows ring orifice increase in solution. *FEBS Lett* 2000; **471**:211–4.
97. Young DB, Garbe TR. Heat shock proteins and antigens of *Mycobacterium tuberculosis*. *Infect Immun* 1991;**59**: 3086–93.
98. Mehra V, Bloom BR, Bajardi AC, et al. A major T-cell antigen of *Mycobacterium leprae* is a 10-kD heat shock cognate protein. *J Exp Med* 1992;**175**:275–84.
99. Cobb AJ, Frothingham R. The GroES antigens of *Mycobacterium avium* and *Mycobacterium paratuberculosis*. *Vet Microbiol* 1999;**67**:31–5.
100. Kim J, Sette A, Rodda S, et al. Determinants of T cell reactivity to the *Mycobacterium leprae* GroES homologue. *J Immunol* 1997;**159**:335–43.
101. Chua-Intra B, Peerapakorn S, Davey N, et al. T-cell recognition of mycobacterial GroES peptides in Thai leprosy patients and contacts. *Infect Immun* 1998;**66**: 4903–9.
102. Meghji S, White PA, Nair SP, et al. *Mycobacterium tuberculosis* chaperonin-10 stimulates bone resorption: a potential contributory factor in pott's disease. *J Exp Med* 1997;**186**:1241–6.
103. Abou-Zeid C, Smith I, Grange JM, Ratliff TL, Steele J, Rook GA. The secreted antigens of *Mycobacterium tuberculosis* and their relationship to those recognized by the available antibodies. *J Gen Microbiol* 1988;**134**:531–8.
104. Malik ZA, Denning, GM, Kusner DJ. Inhibition of $Ca^{2+}$ signaling by *Mycobacterium tuberculosis* is associated with reduced phagosome–lysosome fusion and increased survival within human macrophages. *J Exp Med* 2000;**191**: 287–302.
105. Pohl E, González A, Hermes C, van Silfhout RG. Overview of the tunable beamlines for protein crystallography at the EMBL Hamburg outstation; an analysis of current, future usage and developments. *J Synchr Rad* 2001;**8**:1113–20.
106. Pohl E, Holmes RK, Hol WGJ. Crystal structure of the iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis* shows both metal binding sites fully occupied. *Mol Biol* 1999;**286**:1145–56.
107. Pym AS, Domenech P, Honore N, Song J, Deretic V, Cole ST. Regulation of catalase-peroxidase (KatG) expression, isoniazid sensitivity and virulence by furA of *Mycobacterium tuberculosis*. *Mol Microbiol* 2001;**40**:879–89.
108. Goulding, C, 2002, unpublished results.
109. Grandoni JA, Marta PT, Schloss VT. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. *J Antimicrob Chemother* 1998;**42**:475–82.
110. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 2000; **289**:1546–50.
111. Morth JP, Tucker P, unpublished results.
112. Jungblut PR, Schaible UE, Mollenkopf HJ, Zimny-Arndt U, Raupach B, Mattow J, Halada P, Lamer S, Hagens K, Kaufmann SHE. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol* 1999;**33**:1103–17.
113. Mollenkopf H-J, Jungblut PR, Raupach B, Mattow J, Lamer S, Zimny-Arndt U, Schaible UE, Kaufmann SHE. A dynamic 2D-PAGE database: the mycobacterial proteome via internet. *Electrophoresis* 1999;**20**:2172–80.
114. Porter TD, Coon MJ. Cytochrome P450: multiplicity of isoforms, substrates and catalytic and regulatory mechanisms. *J Biol Chem* 1991;**266**:13469–72.
115. Kunst F, Ogasawara N, Moszer I, et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 1997;**390**:249–56.
116. Munro AW, Lindsay JG. Bacterial cytochromes P450. *Mol Microbiol* 1996;**20**:1115–25.
117. Nelson DR, Koymans L, Kamataki, et al. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 1996; **6**:1–42.
118. Bellamine A, Mangla AT, Nes WD, Waterman MR. Characterization and catalytic properties of the sterol 14α-demethylase from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1999;**96**:8937–42.
119. Podust LM, Poulos TL, Waterman MR. Crystal structure of cytochrome P450 14 alpha-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. *Proc Natl Acad Sci USA* 2001;**98**:3068–73.
120. Souter A, McLean KJ, Smith WE, Munro AW. The genome sequence of *Mycobacterium tuberculosis* reveals cytochromes P450 as novel anti-TB drug targets. *J Chem Technol Biotechnol* 2000;**75**:933–41.
121. Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein folding assay using green fluorescent protein. *Nature Biotechnol* 1998;**17**:691–5.
122. Yang JK, Yoon H-J, Ahn HJ, Lee BI, Cho SH, Waldo GS, Park MS, Suh SW. Crystallization and preliminary X-ray

crystallographic analysis of the Rv2002 gene product from *Mycobacterium tuberculosis*, a beta-ketoacyl carrier protein reductase homologue. *Acta Crystallogr* 2002;**D58**: 303–5.
123. Scorpio A, Zhang Y. Mutations in pncA, agene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nature Med* 1996;**2**:662–7.
124. Rattan A, Kalia A, Ahmad N. Multidrug-resistant *Mycobacterium tuberculosis*: molecular perspectives. *Emerging Infectious Dis* 1998;**4**:195–209.
125. Wilson M, DeRisi J, Kristensen HH, et al. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* 1999;**96**:12833–8.
126. Hube B, Stehr F, Bossenz M, et al. Secreted lipases of *Candida albicans*: cloning, characterisation and expression analysis of a new gene family with at least ten members. *Arch Microbiol* 2000;**174**:362–74.
127. Cole ST. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett* 1999;**452**:7–10.
128. Kouker G, Jaeger K-E. Specific and sensitive plate assay for bacterial lipases. *Environ Microbiol* 1987;**53**:211–3.
129. van Kampen MD, Rosenstein R, Gotz F, Egmond MR. Cloning, purification and characterisation of *Staphylococcus warneri* lipase 2. *Biochim Biophys Acta* 2001;**1544**:229–41.
130. Langfort J, Ploug T, Ihlemann J, et al. Expression of hormone-sensitive lipase and its regulation by adrenaline in skeletal muscle. *Biochem J* 1999;**340**:459–65.
131. Mdluli K, Sherman DR, Hickey MJ, et al. Biochemical and genetic data suggest that InhA is not the primary target for activated isoniazid in *Mycobacterium tuberculosis*. *J Infectious Dis* 1996;**174**:1085–90.
132. Sherman DR, Voskuil M, Schnappinger D, et al. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc Natl Acad Sci USA* 2001; **98**:7534–9.
133. Alland D, Kramnik I, Weisbrod TR, et al. Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1998;**95**:13227–32.
134. Alland D, Steyn AJ, Weisbrod T, Aldrich K, Jacobs Jr WR. Characterization of the *Mycobacterium tuberculosis* iniBAC promoter, a promoter that responds to cell wall biosynthesis inhibition. *J Bacteriol* 2000;**182**:1802–11.