doi:10.1006/jmbi.2001.4461 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **307**, 671–681

JMB

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

Daniel H. Anderson[1]*, Günter Harth[2], Marcus A. Horwitz[2] and David Eisenberg[1]

[1]*UCLA-DOE Laboratory of Structural Biology and Molecular Medicine University of California Los Angeles Box 951570, Los Angeles CA 90095-1570, USA*

[2]*Department of Medicine CHS 37-121, UCLA School of Medicine, 10833 Le Conte Avenue, Los Angeles CA 90095-1688, USA*

**Corresponding author*

The *Mycobacterium tuberculosis* 30 kDa major secretory protein (antigen 85B) is the most abundant protein exported by *M. tuberculosis*, as well as a potent immunoprotective antigen and a leading drug target. A mycolyl transferase of 285 residues, it is closely related to two other mycolyl transferases, each of molecular mass 32 kDa: antigen 85A and antigen 85C. All three catalyze transfer of the fatty acid mycolate from one trehalose monomycolate to another, resulting in trehalose dimycolate and free trehalose, thus helping to build the bacterial cell wall. We have determined two crystal structures of *M. tuberculosis* antigen 85B (ag85B), initially by molecular replacement using antigen 85C as a probe. The apo ag85B model is refined against 1.8 Å data, to an *R*-factor of 0.196 (R_{free} is 0.276), and includes all residues except the N-terminal Phe. The active site immobilizes a molecule of the cryoprotectant 2-methyl-2,4-pentanediol. Crystal growth with addition of trehalose resulted in a second ag85B crystal structure (1.9 Å resolution; *R*-factor is 0.195; R_{free} is 0.285). Trehalose binds in two sites at opposite ends of the active-site cleft. In our proposed mechanism model, the trehalose at the active site Ser126 represents the trehalose liberated by temporary esterification of Ser126, while the other trehalose represents the incoming trehalose monomycolate just prior to swinging over to the first trehalose site to displace the mycolate from its serine ester. Our proposed interfacial mechanism minimizes aqueous exposure of the apolar mycolates. Based on the trehalose-bound structure, we suggest a new class of antituberculous drugs, made by connecting two trehalose molecules by an amphipathic linker.

Keywords: Mycobacterium tuberculosis; 30 kDa major secretory protein; antigen 85B; mycolyl transferase; drug design

Introduction

Mycobacterium tuberculosis, the primary agent of tuberculosis, is responsible for two million deaths annually worldwide, making it the world's most common cause of death from a single infectious agent (Dye *et al.*, 1999). The emergence of new strains of *M. tuberculosis* resistant to the conventional antibiotics used to treat tuberculosis has given urgency to the search for a better vaccine and new drugs against this pathogen (Grosset, 1996).

M. tuberculosis expresses three closely related mycolyl tranferases, also known as antigen 85 proteins (ag85A, ag85C are 32 kDa; 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 (Belisle *et al.*, 1997). A trehalose molecule consists of two D-glucose molecules linked through their O1

Abbreviations used: Ag85A, Ag85B, Ag85C, antigens 85 A, B, and C; Tre, trehalose, α-D-glucopyranosyl α-D-glucopyranoside; MPD, 2-methyl-2,4-pentanediol; BCG, *Mycobacterium bovis*, strain ''bacillus Calmette-Guérin''.

E-mail address of the corresponding author: dha@mbi.ucla.edu

0022-2836/01/020671–11 $35.00/0

atoms. Mycolic acids have two apolar chains: an α-chain of 22-24 carbon atoms, and a β-chain of typical length 40-60 carbon atoms, but some as long as 120 carbon atoms (Brennan & Nikaido, 1995). Modifications of the β-chain include one hydroxyl group, cyclopropane rings, double bonds, one keto group and terminal carboxyl or methoxy groups. The three antigen 85 variants are expressed at a steady-state ratio of 3:2:1 (ag85B: ag85A:ag85C), although the three genes are unlinked (Harth *et al*., 1996). The 30 kDa antigen 85B is the most abundant of the three, and is the most abundant extracellular protein of *M. tuberculosis*, responsible for nearly one-quarter of the total extracellular protein in broth culture (Harth *et al*., 1996). However, it is the least active by one *in vitro* assay (Belisle *et al*., 1997). The three mycolyl transferases are also abundantly expressed in infected human macrophages (Lee & Horwitz, 1995). In *M. tuberculosis* phagosomes within human macrophages, the antigen 85B molecules are found in the phagosomal space and on the bacterial cell wall, as demonstrated by the cryosection immunogold technique (Harth *et al*., 1996).

The 30 kDa major secretory protein of *M. tuberculosis* (antigen 85B) is a leading vaccine candidate (Horwitz *et al*., 1995). Vaccination of guinea pigs with the purified *M. tuberculosis* ag85B induces substantial protective immunity against aerosol challenge with *M. tuberculosis* bacteria (Horwitz *et al*., 1995). Vaccination of guinea pigs with recombinant *Mycobacterium bovis* BCG expressing and secreting the *M. tuberculosis* 30 kDa protein induces stronger protective immunity against aerosol challenge than conventional *M. bovis* BCG vaccine (Horwitz *et al*., 2000a,b). This recombinant BCG vaccine is the first vaccine demonstrated to be more potent than conventional BCG vaccine since the latter was developed nearly a century ago.

Because of its location at the cell wall and its involvement in cell wall biogenesis, the antigen 85B protein of *M. tuberculosis* may be a relatively accessible drug target, an hypothesis based upon studies of another *M. tuberculosis* major extracellular protein, glutamine synthetase. Glutamine synthetase is thought to be involved in the synthesis of a poly-L-glutamate/glutamine cell wall structure found exclusively in pathogenic mycobacteria. A compound that inhibits *M. tuberculosis* glutamine synthetase has been shown to block the synthesis of this cell wall structure and to inhibit *M. tuberculosis* growth in broth culture, in infected human macrophages, and in infected guinea pigs (Harth & Horwitz, 1999; G.H. & M.A.H., unpublished data). By analogy, inhibitors of the enzymatic activities of other *M. tuberculosis* secreted proteins may also arrest bacterial growth. Also consistent with the idea that the 30/32 kDa complex is a promising drug target, a 28-mer antisense oligodeoxyribonucleotide fully complementary to the 32A kDa protein (ag85A) mRNA and mismatched at only one nucleotide with the mRNA of the 30 and 32B kDa proteins (ag85B and ag85C, respectively) inhibits *M. tuberculosis* growth in broth culture (G.H., P.C. Zamecnik & M.A.H., unpublished data). Deletion of the gene encoding the antigen 85C protein alters the bacterial cell wall and its permeability, but does not kill the cells (Jackson *et al*., 1999). Although some mutations in ag85A inhibit cell growth to some extent (Armitige *et al*., 2000), it may be necessary to inhibit all three antigen 85 variants to arrest cell growth. It is pertinent that even a bacteriostatic inhibitor can halt disease progression. Several common antituberculous drugs are bacteriostatic but not bactericidal against *M. tuberculosis*, including para-amino salicylic acid, ethionamide, and thiacetazone (Kucers & Bennett, 1987). Mycolyl transferase activity is unique to mycobacteria; there is little chance that an inhibitor of ag85B would accidentally inhibit a human enzyme. Antigen 85B is a candidate as a vaccine and as a drug target.

Crystal structures of apo and diethylphosphate-inhibited *M. tuberculosis* antigen 85C were previously determined by Ronning *et al*. (2000), who also proposed a reaction mechanism. The ag85C has an α-β hydrolase fold (Nardini & Dijkstra, 1999) with a Ser-His-Glu catalytic triad. Reaction proceeds through an intermediate mycolate-esterified Ser124. The mycolyl transferase process occurs in a mostly hydrophobic groove. Prior to forming the covalent acyl intermediate, the mycolate α-chain could insert into a tunnel adjacent to the active serine residue.

We report herein the molecular replacement solution of crystal structures of the *M. tuberculosis* antigen 85B with and without bound trehalose. We discuss two principal insights gained from our analyses of ag85B structural features. We propose a refinement of the basic reaction scheme previously proposed for ag85C (Ronning *et al*., 2000): an interfacial catalysis process consistent with the association of this soluble enzyme with the bacterial surface. We extrapolate from the ligands observed in the antigen 85B active-site groove to a new class of inhibitors for the three antigen 85 mycolyl transferases.

Results

Comparison of the mycolate transferases ag85B and ag85C

As expected from the 73% amino acid sequence identity, the ag85B and ag85C structures are similar. The ribbon representation in Figure 1 summarizes their α-β hydrolase fold (Nardini & Dijkstra, 1999). For superposition of apo ag85C onto apo ag85B, LSQKAB (see Materials and Methods) fit all backbone atoms of residues 7-29, 33-84, and 96-283 (in ag85B numbering); regions that differ greatly (2-6, 30-32, 85-95, and 284-5) were excluded from the calculation. The rms distance between the

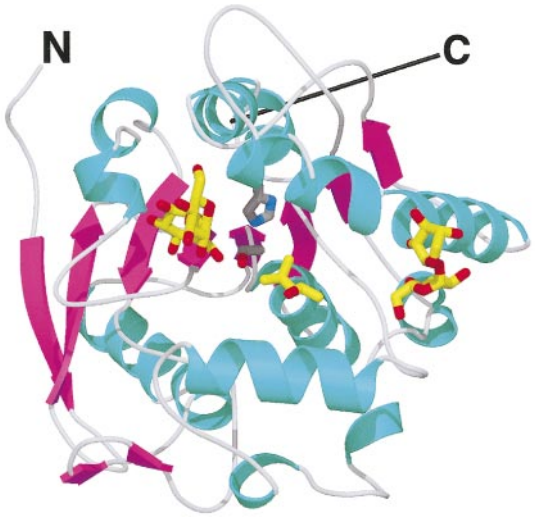

Figure 1. Ribbon diagram of trehalose-bound ag85B, showing its α-β hydrolase fold. Molscript (Kraulis, 1991) automatically assigned the secondary structural elements by its backbone φ and ψ criteria, and produced most of this Figure. Cyan ribbons represent α-helices, magenta ribbons β-strands, coil and turn regions are white. The N terminus is at upper left, and the C terminus is at the back of the molecule. Active site Ser126 and His262 CA and side-chain atoms are shown (protein C atoms are gray; N, blue; O, red). Trehalose 1151 is on the left, Trehalose 1152 is on the far right, and MPD 1001 is at the center (all with yellow C atoms and red O atoms). The MPD at this site may be a surrogate for the mycolate ester intermediate. Figures 1, 3 and 5 were produced using XtalView (McRee, 1993), and were rendered with Raster3D (Merritt & Bacon, 1997).

selected backbone atom pairs is 0.50 Å. The structures diverge at the two-residue insertion around residue 30, and at the N and C termini. The loop of residues 85-95 influenced by the Cys87 to Cys92 disulfide bridge is very different from the corresponding unbridged loop in ag85C, with some corresponding backbone atoms 5-6 Å apart. In the ag85C sequence, Gln85 and Asn89 correspond to the bridged cysteine residues in ag85B. Much of the geometry is preserved within corresponding secondary structural elements of both molecules, but several elements are slightly tilted or shifted relative to their counterparts, as was observed during initial model rebuilding. In the core β-sheet, most of the side-chain rotamers match, and the few sequence substitutions are complementarily paired to maintain total volume. The depth of the active site groove is about the same in antigens 85B and -C, but the ag85C active site groove seems more closed than the ag85B groove. Adjacent to the catalytic Ser126, Leu152 in ag85B replaces Phe150 in ag85C. The adjacent change from Trp158 (ag85C) to Gly160 (ag85B) allows a rotamer change at position 152 that preserves similar active site volumes.

It is not immediately obvious from the structural differences discussed above if antigens 85B and 85C play distinct biological roles, or why *M. tuberculosis* expresses three antigen 85 variants. The only *in vitro* assay so far (Belisle *et al*., 1997) found that ag85C is about eight times more active than ag85B, but did not explore substrate specificities. An amphiphilic molecule of 2-methyl-2,4-pentanediol (MPD) binds to the active site of ag85B in our study, but was not reported by Ronning *et al*. (2000) in their study of ag85C using the same cryoprotectant. This could be evidence that the two enzymes prefer different oxygenation patterns in their substrates.

Trehalose binding

Antigen 85B binds trehalose in two sites. Discovery and modeling of the two trehalose molecules is discussed in Interpretation of non-protein density, below. We present our interpretation and use of the two sites in Proposed reaction process, and Proposed drug design, in the Discussion below. The two bound trehalose molecules contact the ag85B enzyme directly and through intermediary water molecules (Figure 2). Contacts to protein backbone atoms are more fold-dependent than sequence-dependent. One trehalose molecule (Tre1151; on the left in Figure 1) binds adjacent to the catalytic Ser126, with some contacts as predicted by Ronning *et al*. (2000). Its binding displaces five water molecules from the apo ag85B structure, and adds three new water molecules (for details, see the legend to Figure 2). Binding of Tre1151 buries 245 Å^2 of enzyme surface. The other trehalose molecule (Tre1152; on the right in Figure 1) binds at the other end of the active-site groove. Trehalose binding at the non-catalytic Tre1152 site displaces only two water molecules from the apo structure, binds one new water molecule, and buries 192 Å^2 of enzyme surface. As a consequence of its greater interactions with the enzyme, the electron density is better for Tre1151 than for Tre1152 (Figure 3). By visual inspection of the model, it appears that enough space surrounds Tre1152 atom C4P (top of upper ring of Figure 2(b)) to accommodate a galactose ring (inverted configuration at C4P). Glucose-galactose linkages have not been observed in mycobacteria. In Proposed reaction process below, we suggest a functional role for the wobbly ring of Tre1152.

Protein atoms near and far from the trehalose molecules rearrange on addition of trehalose. The rms difference for all backbone atoms in residues 2-285 is 0.23 Å, for all protein atoms 0.44 Å. The ring of Phe232 swings out of the way of Tre1152, while Met159 becomes more ordered by atom CE contacting the sugar surface. The apo ag85B electron density map suggests partial occupancy of a

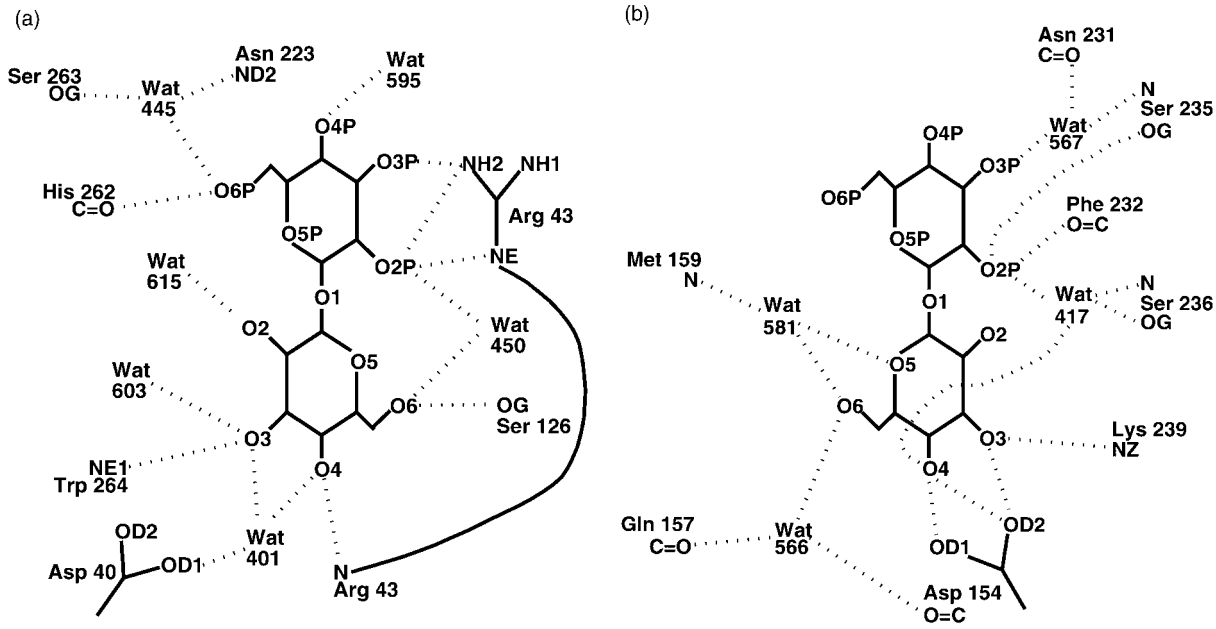

Figure 2. Schematic diagrams of trehalose hydrogen bonds to ag85B protein. Hydrogen bond contacts were determined automatically with XtalView (McRee, 1993). (a) Trehalose 1151 binds at the active site Ser126, has many contacts, and is well ordered in the crystal. Wat401 was the most prominent non-protein density in the apo ag85B maps. Nearby hydrophobic atoms are: Leu42 CB, Pro225 CB, Leu229 CD1, catalytic triad His262 CD2 and Trp267 CH2 and CZ3. Trehalose binding at this site displaced five water molecules from the apo ag85B structure: water molecules 418, 442, 455, 528, 574. In addition, Wat567 binds to Arg43 in the apo structure in a site not overlapping Tre1151, but is absent from the trehalose model. Water molecules 595, 603 and 615 bind along with Tre1151. (b) Trehalose 1152 binds at the other end of the active site groove. In the text and Figure 4 we propose that this second site is the second trehalose monomycolate substrate-binding site. Fewer protein contacts result in more disorder and worse density for the upper (primed) ring (Figure 3(c)). Nearby hydrophobic atoms are: Met159 CE (disordered in apo ag85B), Phe232 ring, CD1 atoms of leucine residues 152 and 163. Binding trehalose at the non-catalytic Tre1152 site displaces water molecules 460 and 543 of the apo ag85B structure, and adds the new Wat567.

second conformation for Phe232, close to its average position in the trehalose structure. A few sidechains far from the trehalose molecules were beyond automatic capture by SHELXL. For example, about 180° torsion changes are seen at Gln107 (χ_1), Leu219 and Leu238 (χ_2, with smaller accomodating χ_1 torsions). These sites appear to have little rotamer selectivity.

Discussion

Structure-based insight into the reaction mechanism

Based on our structures of ag85B, we are able to add detail to the mycolyl transferase reaction mechanism proposed on the basis of the ag85C structures (Ronning *et al.*, 2000). The ag85B and ag85C structures have largely apolar, curved tunnels leading away from the active site serine (about 15 Å along a line from opening to opening; see Figure 2 in the report by Ronning *et al.*, 2000). Ronning *et al.* (2000) suggest that this would be a suitable binding site for the α-chain of the mycolate. The hydrophobic active-site groove and adjacent exterior surfaces (see below) could funnel the first mycolate α-chain into the tunnel, an otherwise improbable event. In our apo and trehalose ag85B models, six water molecules bind in the tunnel and would have to be displaced by that α-chain. The surface opening of this tunnel could accommodate varying α-chain lengths, and allow water extrusion. Mycolates in *M. tuberculosis* have α-chain lengths of 22-24 carbon atoms. Alternatively, the tunnel could also function as a H^+ conductance channel for the possibly sealed active site (see below).

Proposed reaction process

We propose a reaction process inferred from these ag85B crystal structures, and from some previously published experimental observations. Our proposed moving-enzyme process is complementary to the chemical mechanism proposed by Ronning *et al.* (2000). The schematic in Figure 4 illustrates our proposed major events of catalysis.

In the trehalose-bound ag85B crystal structure, one O6 atom of Tre1151 is near Ser126 OG, as it

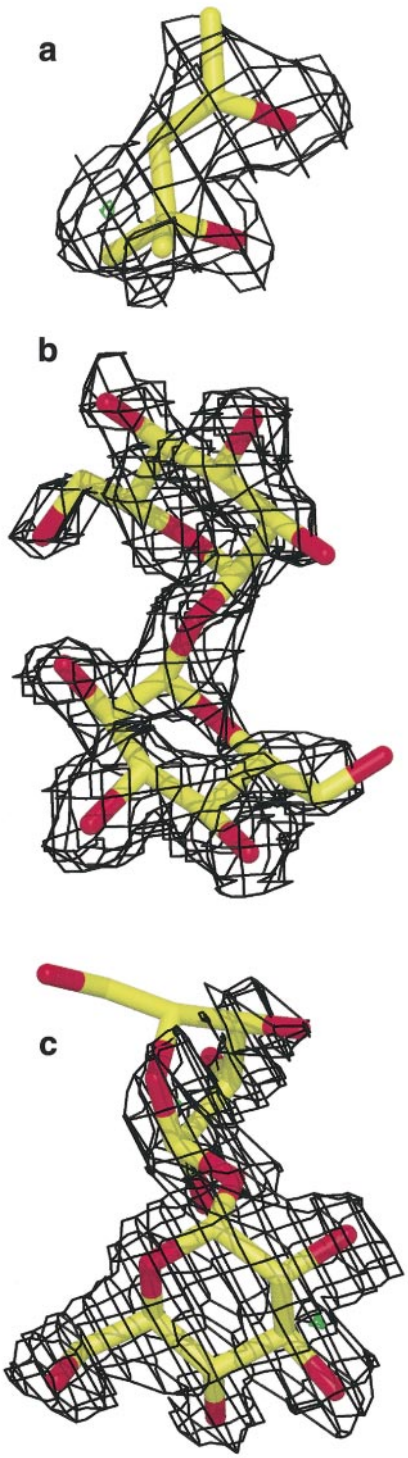

Figure 3. Atomic models and electron densities representing binders to the active groove of ag85B. C atoms are yellow, O atoms are red. The black density maps have Fourier coefficients $2F_{obs} - F_{calc}$, final refined model phases, include all measured data, and are contoured at 1σ. No significant peaks appear in these regions of the $F_{obs} - F_{calc}$ maps (contoured at ±3σ; green and red), meaning that the models mimic the crystalline atoms well. (a) MPD1001 at the active site of the apo ag85B structure (data resolution range is 20-1.8 Å). This MPD could mimic the head region of a mycolate covalently bound to Ser126. (b) Trehalose 1151 density (data resolution range is 20-1.9 Å). Both trehaloses are oriented to approximate the flattened rendition in the schematics of Figure 2, each with its primed ring at top (atom names C4P, O2P, etc.). Trehalose 1151 density shows a hole in each ring, meaning that it is well clamped. Atom O6 (lower right) is near the catalytic Ser126, but its weak density could indicate a low hydrogen bonding energy. (c) Trehalose 1152 binds at the other end of the active-site groove. Density for the lower ring has a small hole. The upper (primed) ring is less confined by protein atoms (Figure 2(b)), and therefore atoms C5P, C6P, O6P (upper left) produce no density at this contour level. Electron density vanishes for trehalose atoms whose *B*-factors are over about 80 Å^2.

would be immediately after forming the covalent serine mycolate ester intermediate. In the reaction intermediate, the atoms in the ester bond would push the trehalose away from its location in the crystal structure. Alternatively, the nucleophilic trehalose O6 atom would reverse the partial reaction by displacing the mycolyl ester from Ser126.

The molecule of the cryoprotectant 2-methyl-2,4-pentanediol bound adjacent to the active site Ser126 (MPD1001; location shown in middle of Figure 1, density in Figure 3(a)) could represent the oxygenated head region of a bound mycolate. MPD1001 binds in the hydrophobic groove across the enzyme, on surfaces contiguous with the hydrophobic ''collar'' discussed below. This is the same groove that Ronning *et al*. (2000) suggest to be the site of the mycolate β-chain. We suggest that the β-chain of the intermediate drapes across this hydrophobic surface and towards the cell wall as part of the anchor, not across the groove.

The other trehalose molecule (Tre1152) binds adjacent to the same hydrophobic surface and groove as Tre1151. Trehalose 1152 could represent part of a trehalose monomycolate tethered to the cell surface. One Tre1152 ring protrudes from the enzyme and could point its atom O6P toward the cell. The other Tre1152 ring is more caged by protein atoms (Figure 2(b)); protein atoms crowd its atom O6 too much for it to hold a mycolate chain. Phe232 rotates mostly around χ_1 on binding trehalose, and is surrounded by sufficient space to rotate further and to function as a hydrophobic gate over the active site groove. A trehalose monomycolate binding at the site of Tre1152 could swing its mycolate α-chain over the enzyme surface and into the groove when Phe232 rotates. We propose that Tre1151 represents the departing trehalose leaving behind mycolate ester, MPD 1001 represents the mycolate ester intermediate, and Tre1152 represents the incoming trehalose monomycolate (see Figure 4).

The exposed hydrophobic surfaces adjacent to the active-site groove of ag85B suggest interfacial catalysis, and fluidity over the cell surface as part of the enzyme's function, analogous to the ''scooting mode'' interfacial catalysis of phospholipases A_2 (Gelb *et al*., 1999). Phospholipases A_2 appear to anchor to their aggregated substrates *via* charges and exposed tryptophan residues, and ''seal'' their active-site grooves with surrounding hydrophobic ''collars''. The antigen 85B collar (Figure 5(a)) has one salt bridge, Arg43 to Asp170, and hydrophobic residues Met159 (Trp157 in ag85C), Leu163, Leu166, Ile224, Pro225, Phe228, Leu229 and Phe232 (Leu230 in ag85C).

In summary, secreted ag85B localizes to the mycobacterial cell wall (Harth *et al*., 1996). In the trehalose-bound ag85B crystal structure we find trehalose molecules in two binding sites. One trehalose is oriented to point a mycolate at the cell wall. A molecule of MPD marks the start of the mycolate intermediate, anchored to the cell wall. At what would then be the ag85B surface nearest the cell wall we find a ring of hydrophobic residues analogous to phospholipase A_2 collar residues. Based on this accumulated evidence, we propose that antigen 85B operates in constant contact with the mycobacterial cell (Figure 4). We further propose that the second trehalose binding site helps anchor the ag85B enzyme to the cell surface, while selecting for the exposed O6 atom of the next trehalose monomycolate substrate. Allowing one wobbly trehalose ring at the non-catalytic site increases the search radius. Thus, the non-catalytic trehalose binding site can confer some directionality to the diffusion of the enzyme across the cell surface, enhancing catalysis.

Our proposed interfacial reaction process minimizes aqueous exposure for the extremely hydrophobic mycolate chains. The hydrophobic groove, possibly sealed by its surrounding collar, functions as a small volume of organic solvent migrating across the cell wall. The apparent driving force for the transferase process is that it is much less likely that a trehalose dimycolate will extrude partway out of the cell wall to contact Ser126 and reverse the tranferase reaction, than for trehalose monomycolates to sequentially slide into the active site groove. In our interfacial reaction scenario, the putative fibronectin-binding surface is left exposed (Naito *et al*., 1998; Ronning *et al*., 2000). Thus, our analysis of structural features yields insight into the function of antigen 85B.

Proposed drug design

We propose that a class of groove-binding inhibitors of the antigens 85 would result from linking two trehalose molecules by the proper length amphiphilic chain. Figure 5(b) shows the imagined inhibitor, extrapolated from the ligands found by our crystallographic experiment. Hydrophilic linker atoms would face in or out of the groove as appropriate. For example, MPD 1001 in the middle of Figure 5(a) became part of the linker in Figure 5(b). The accumulation of contacts could result in tight binding that gags the enzyme's active-site groove. The molecular mass of the prototype inhibitor shown in Figure 5(b) is 963, in the range of ingestible drugs. On the way to the target cell, the hydrophobic surfaces of the linker could be sequestered by coiling between the solubilizing trehalose portions, whose polar surface character might aid in biodelivery. The three antigen 85 variants might require three drug variants for full inhibition of their mycolyl transferase activities. Such drugs could compete for contacts between the antigens 85 and the cell wall, inhibiting growth and disease progression.

Materials and Methods

Purification

M. tuberculosis antigen 85B protein (ag85B) was expressed in *Mycobacterium smegmatis* and purified as described (Harth *et al*., 1997). The ag85B was further purified by elution from silica-based size-exclusion columns (TosoHaas G3000SW preparative or G3000SWxl analytical). The elution buffer was 0.025 M sodium phosphate, 0.1 M Na_2SO_4, 1 mM NaN_3, titrated together to pH 6.5. The ag85B binds to the column, and elutes in two very late peaks. The minor first peak contains ag85B missing 148 mass units (measured by electrospray mass spectrometry), as though the N-terminal Phe residue is missing. The major second peak (eluting from the preparative column after one hour 45 minutes at 3 ml/minute) was concentrated in Centricon 10 followed by Microcon 10 centrifugal devices (Millipore/Amicon). The second peak contains full-size ag85B, and was used for crystallization, but the crystallographic experiment did not localize the N-terminal Phe residue. Both peaks contain monomeric ag85B, as measured by light scatter from the analytical column eluent (Wyatt Technology 3-scatter-angle miniDAWN).

Crystallization

We were unable to improve the growth of Form 1 crystals. A typical reservoir solution resulting in Form 1 clusters of plates was 0.35 M LiCl, 0.05 M lithium citrate, pH about 3.5. The hanging drops contained equal volumes of ag85B and reservoir. In retrospect, a crystallization response surface with zero derivatives cannot be optimized (Carter, 1997). Form 2 crystals and thus the structures reported here resulted from skipping elsewhere in experimental parameter space.

Form 2 crystals were grown by hanging-drop vapor diffusion. The ag85B, still dissolved in the above elution buffer, was re-diluted with elution buffer to about 16 mg/ml, and then mixed with an equal volume of reservoir solution. Enough volume of this drop mixture to prepare four or six drops was centrifuged (about 10,000 *g* for about five minutes), and then 5 µl drops were hung over 1 ml reservoirs containing: 0.6-0.7 M $(NH_4)_2SO_4$, 22.5 % 2-methyl-2,4-pentanediol (MPD), 0.05 M Na 2-(*N*-morpholino)ethanesulfonate (MES), pH 6.0. Trigonal prism crystals grew in a few days. Crystal-containing drops were flooded with reservoir to prevent phase separation while handling the crystals for flash-cooling in a cryostream (Parkin & Hope, 1998). The diffraction quality frequently improved with annealing (Harp *et al*., 1999).

An additional crystal was grown as above, except 37 % trehalose solution (nearly saturated) instead of water was used to maintain a constant 1 ml volume in the reservoir. The reservoir was 0.6 M $(NH_4)_2SO_4$, and therefore the trehalose concentration was about 0.53 M. This crystal was flash-cooled by immersion in liquid nitrogen, annealed in room temperature reservoir solution, then re-frozen in a cryostream.

Structure determination and evaluation

The apo ag85B structure in Form 2 crystals was determined by molecular replacement using AMoRe (Navaza, 1994) as implemented in CCP4 (CCP4, 1994). The search fragment was molecule A of apo ag85C (Ronning *et al*., 2000), with N-terminal residues 3-5 deleted because they visibly differ between molecules A and B of ag85C. We used a room temperature diffraction data set. All diffraction data used in this work were processed with DENZO and SCALEPACK (Otwinowski & Minor, 1997). We expected the ag85B and ag85C structures to be very similar (sequence identity is 73 %). We therefore used approximate *E* values rather than *F* values in AMoRe to intensify the peaks. The rotation function yielded a single peak nearly twice as large as the second peak (integration radius set to 20 Å). Similarly the two translation function peaks were nearly identical, with very little background.

Rigid body refinement in AMoRe was followed by further rigid body refinement with REFMAC (CCP4, 1994; Murshudov *et al*., 1997). Model-phased maps calculated at this point (with coefficients $2F_{obs} - F_{calc}$ and $F_{obs} - F_{calc}$) illustrated the sequence changes and regions needing retracing. Thus we began rebuilding the model alternately using XtalView (McRee, 1993) for display, and REFMAC for refinement. As better diffraction data were obtained from frozen crystals, we kept the same model-naïve test set reflections (Brünger, 1997) chosen by the CCP4 ''uniqueify'' mechanism (CCP4, 1994). When the model was nearly complete, we switched to refinement with SHELXL (Sheldrick & Schneider, 1997). Addition of ''riding'' hydrogens to the apo ag85B model improved R_{free} by some combination of the added anti-bump restraint and the additional hydrogen scattering density. The apo and trehalose ag85B crystals were nearly isomorphous (scaling *R* was 0.18). The trehalose structure determination began by refinement of the apo model against trehalose diffraction data using SHELXL. Both models include bulk solvent correction and anisotropic scaling. Both final ag85B models have been refined against diffraction data obtained from cryoprotected crystals at synchrotrons (see Acknowledgments). For the first and most recent data sets, we used Quad4 CCD detectors (ADSC).

Four asparagine side-chain orientations were assigned so as to position their slightly positive ND2 atoms adjacent to the slightly negative NE atoms of tryptophans (Souhassou *et al*., 1991). This charge-density based assignment simultaneously optimized the hydrogen bonds for these asparagine residues.

The final apo ag85B model (without trehalose) contains all residues but the N-terminal Phe residue. Electron density is very poor for Met159, some exposed side-chains, and the C-terminal Gly285. Density for Met159 improved as a result of binding trehalose, but the displacement parameter of atom SD is still very high in the final trehalose model.

The geometries of the apo and trehalose ag85B models, as evaluated by Ramachandran plots (Ramachandran *et al*., 1963), are almost entirely in most favored regions. The residues that deviate are located at transitions between secondary structural elements. The active site Ser126 backbone ($\phi = 63.0^\circ$, $\psi = -129.5^\circ$ in apo form) is severely strained, as it is in ag85C and other proteins of the same fold (Nardini & Dijkstra, 1999; Ronning *et al*., 2000). Asn217 ($\phi = -117.8^\circ$, $\psi = -158.7^\circ$ in apo form) hydrogen bonds to a backbone carbonyl group on the opposite side of a loop, and is thus subject to forces not considered in drawing the Ramachandran most favored regions. In the final models, the peptide bond between Gln25 and Phe26 is not flat, despite the flatness

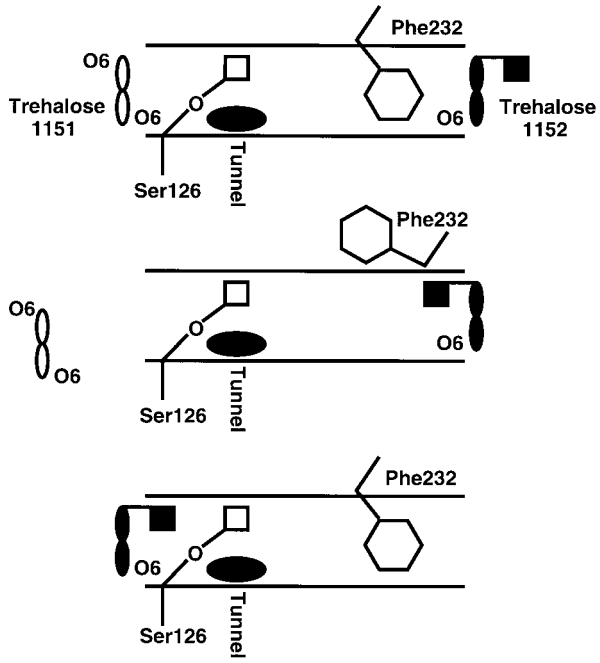

Figure 4. Schematic diagram of proposed mycolyl transferase mechanism. This "scooting" mechanism (Gelb *et al.*, 1999) is an elaboration on the proposed chemical mechanism (Ronning *et al.*, 2000). Pairs of horizontal lines represent the active-site groove. The narrow double ellipses represent the rings of the bound trehalose molecules. "O6" represents a free trehalose O6 hydroxyl. The squares represent mycolates. The top diagram shows the product of the first step: one trehalose (at the site of Tre1151) has already transesterified its mycolate (open square) onto the active site Ser126. The mycolate α-chain could thread into the tunnel (filled ellipse), while the β-chain extends towards the cell wall (out of the paper). We propose that Tre1152 (filled double ellipse at far right) represents the incoming trehalose monomycolate that will displace the mycolate ester from Ser126. One Tre1152 O6 atom (protruding towards the viewer) begins the mycolate tether (filled black square) to the cell surface. The middle panel shows the second step: A χ_1 rotation of Phe232 (sidechain swings up in this view) opens the lid over the hydrophobic groove between the Tre1152 site and Ser126, and the groove stabilizes the mycolate α-chain outside of the cell wall (shown by the filled black square flipped into the groove). Meanwhile, the free trehalose at left vacates the active site. The bottom panel shows the status midway through the final step: The α-chain in the groove provides a guide and tether for flipping the trehalose monomycolate (Tre1152) over to the Tre1151 site (the black square and double ellipse are now at left). In the reference frame of the tuberculosis cell, the enzyme guided by the α-chain "scoots" across the cell wall to position the second trehalose monomycolate at the active site. The free trehalose O6 atom will then displace the first mycolate (open square) from its ester bond with Ser126. Thus, *M. tuberculosis* exploits the symmetry of trehalose. A wall-tethered mechanism similar to our proposal would accomplish the transfer of the extremely hydrophobic mycolates with minimal intermediate aqueous exposure. Back reaction is likely prevented by the improbability of extracting a trehalose dimycolate out of the cell wall and into the enzyme's groove.

restraint applied throughout the refinement process. WHAT_CHECK (Hooft *et al.*, 1996) detects a few unsatisfied hydrogen bonds. The ag85B crystallizes despite one unfavorable lattice contact (Gly5 backbone O is 2.42 Å from Glu227 OE2 of an adjacent molecule).

The crystallographic *R*-factor for the final apo ag85B model is 0.196 (including all data 20-1.8 Å). For the final refinement with the test set excluded, R_{free} is 0.276. The final trehalose model results in an *R*-factor of 0.195 (all data 20-1.9 Å); R_{free} is 0.285 in the final refinement with excluded test set. In this case, the relatively large differences between *R*-factor and R_{free} (Brünger, 1997; Tickle *et al.*, 1998) we attribute to an inexact space group symmetry. Freshly grown and frozen apo ag85B crystals have trigonal diffraction geometry, but when left at room temperature for several weeks evolve to apparent *P*1 symmetry. The symmetry of the trigonal form is $P3_121$, with cell parameters $a = 73.2$ Å, $c = 92.5$ Å. Typical *P*1 cell parameters are: $a = 73.4$ Å, $b = 74.4$ Å, $c = 92.1$ Å, $\alpha = 89.4°$, $\beta = 89.9°$, $\gamma = 119.6°$. The trehalose-containing crystal was also frozen soon after it was observed. With both crystals, we truncated the data resolution by the criterion of high *R*-factor, not low average I/σ (Apo ag85B data: $R_{sym} = 0.112$ overall, 0.186 in the highest resolution shell with $I/\sigma = 6.0$; Trehalose data: $R_{sym} = 0.068$ overall, 0.367 at high resolution cutoff with $I/\sigma = 7.7$).

To compare structures, we superimposed coordinates with program LSQKAB from the CCP4 suite (CCP4, 1994; Kabsch, 1976), with input atoms as specified in Results. Ligand contact surface areas were calculated with CCP4 program AREAIMOL (Lee & Richards, 1971), using a probe radius of 1.4 Å.

Interpretation of non-protein density

We modeled most of the non-protein electron density as water molecules, but we identified some density as other solvent species. The final apo ag85B model contains 205 bound water molecules, the final trehalose model 223. The apo maps indicated further hydration, including some clathrates over hydrophobic atoms. Adding more waters to the apo model, even with occupancies set to 0.5, reduced the *R*-factor while increasing R_{free} by an almost equal amount. At the start of rebuilding from the ag85C to the ag85B structure, a region of connected density appeared in the active site. This density was left unmodeled until the switch to SHELXL clarified its shape. We fit a molecule of the cryoprotectant 2-methyl-2,4-pentanediol into this density (MPD1001; Figure 3(a)). Two other density features we also interpreted as MPD. MPD1001 and 1003 are convincing explanations of their density in the apo ag85B map. Our fit of MPD1002 is more speculative; its density is flat, as though representing a superposition of MPD molecules in various orientations. In the trehalose-bound map, the shape of the MPD1001 density is less distinctive than in the apo ag85B map, as though it became less tightly bound. The MPD1001 isomer is the same in apo and trehalose models, but our assignment of the rotamer of the dimethyl portion is different. In both structures, the large end of MPD1003 is clasped between Gln141 and His139, but the small end is not well enough restrained to produce good density in the trehalose map. A curved region of density in the apo map we interpreted as a loosely bound molecule of MES buffer. The sulfonate portion is adjacent to the guanidinium of

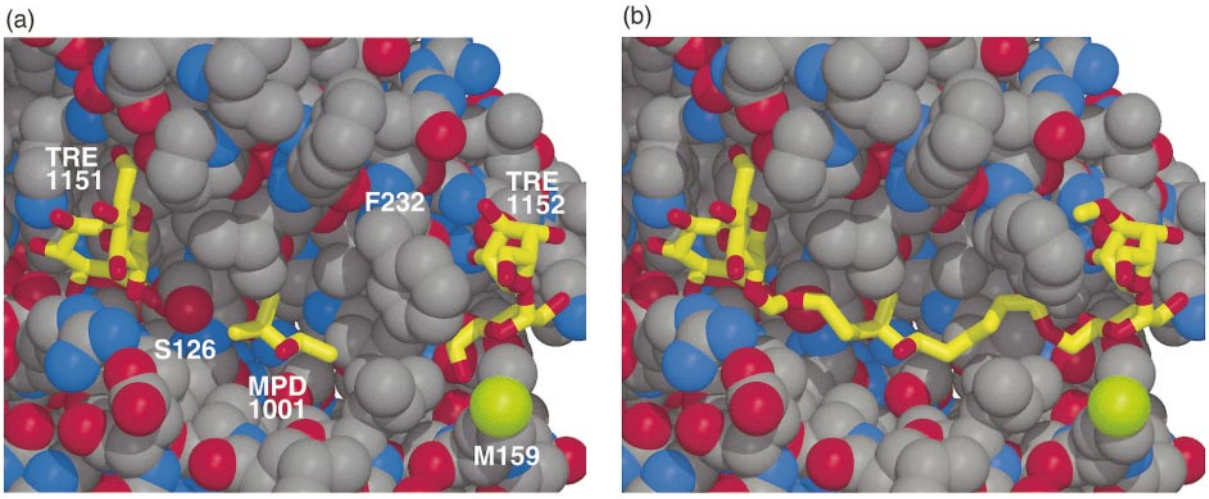

Figure 5. Active site hydrophobic cleft in ag85B, and a hypothetical drug. (a) Protein atoms are shown as balls (drawn at 75 % of van der Waals radii; gray for carbon; blue for nitrogen; red for oxygen; and green for sulfur). The foreground gray atoms are the hydrophobic collar that we propose helps adhere the enzyme to the cell wall (see Proposed reaction process). Hydrophilic surface begins at lower left. Trehalose and MPD molecules are shown as stick figures (yellow carbon atoms and red oxygen atoms). Tre1151 is at far left, with one O6 atom in a hydrogen bond to the active site Ser126 (in the shade just right of Tre1151). The Tre1151 site is where the first trehalose would be immediately after transfer of its mycolate to a temporary ester bond on Ser126 OG. MPD1001 is near the center, also contacting Ser126. The MPD could mimic the partially hydrophilic head of the intermediate mycolate ester. The mycolate α-chain could insert into the tunnel as proposed by Ronning *et al*. (2000) (behind and below the MPD; not fully visible in this view). We propose that the mycolate β-chain protrudes towards the tuberculosis cell wall (towards the viewer). Binding of Tre1152 (far right) has displaced the adjacent Phe232, and partially ordered Met159 (lower right). (b) The proposed class of drugs consists of two trehalose molecules with amphiphilic linkers. The illustrative linker in this Figure is anchored at atom O2P of Tre1151 (at left; atom O6 is too crowded; see Figure 2(a)). To prepare the illustration in (b), we threaded the linker through Wat450 and MPD1001. A branch near the MPD1001 site could plug the tunnel; some of the central linker atoms could mimic a mycolate. As discussed in Proposed reaction process, this Figure shows Phe232 torsioned out of the groove. Phe232 could be encouraged into this new location by a hydrophobic group attached to atom O3P of Tre1152 (at right; here shown as an added methyl). This imaginary linker terminates at atom O6 of Tre1152; the C5-C6 bond is rotated to maintain reasonable geometry.

Arg206, while the ring nitrogen atom we think is likely protonated because of a potential hydrogen bond to OD1 of Asn251. The MES density is worse in the trehalose map.

The locations of the two bound trehalose molecules were first determined by a difference Fourier summation with coefficients ($F_{\text{trehalose}} - F_{\text{apo}}$) and phases from apo ag85B. Refinement of the apo model into the trehalose data improved the phases enough to place Tre1151 easily, and Tre1152 with less ease. One ring of Tre1152 is less caged by protein atoms than the other three sugar rings. Its mobility results in no density above 1σ contour level for atoms C5P, C6P and O6P, even with refined phases. Figure 3 shows trehalose electron density in the final map, and Figure 2 shows trehalose contacts. The ribbon diagram (Figure 1) shows the trehalose locations relative to the fold of the enzyme.

The MPD and MES model geometries were tightly restrained to match the database models. The trehalose restraint stringencies were SHELXL defaults, generated by SHELXPRO.

Protein Data Bank accession codes

MPD and MES coordinates were obtained from the HIC-up database (Kleywegt & Jones, 1998), and were derived from PDB entries 3AL1 and 3CHB, respectively. Trehalose coordinates were first obtained from HIC-up (modified from PDB code 1BYK), then replaced by coordinates from the Cambridge Crystal Structure Database (structure code DEKYEX; Allen *et al*., 1991).

The antigen 85B models and diffraction data have been deposited with the PDB, under accession codes 1F0N for apo, and 1F0P for trehalose.

Acknowledgments

We thank Donald Ronning and James Sacchettini for generously providing ag85C coordinates prior to publication. We also thank Duilio Cascio, Barbara Jane Dillon, Kym Faull, and Cameron Mura.

Diffraction intensities were measured at Brookhaven National Laboratory beamline X8C, and Advanced Light Source (Berkeley) beamline 5.0.2. This work was supported by NIH grants to M.A.H. and D.E.

References

Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). The development of versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **31**, 187-204.

Armitige, L. Y., Jagannath, C., Wanger, A. R. & Norris, S. J. (2000). Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and in macrophages. *Infect. Immun.* **68**, 767-778.

Belisle, J. T., Vissa, V. D., Sievert, T., Takayama, K., Brennan, P. J. & Besra, G. S. (1997). Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**, 1420-1422.

Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu. Rev. Biochem.* **64**, 29-63.

Brünger, A. T. (1997). Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366-396.

Carter, C. W. (1997). Response surface methods for optimizing and improving reproducibility of crystal growth. *Methods Enzymol.* **276**, 74-99.

Collaborative Computing Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999). Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *J. Am. Med. Ass.* **282**, 677-686.

Gelb, M. H., Cho, W. & Wilton, D. C. (1999). Interfacial binding of secreted phospholipases A_2: more than electrostatics and a major role for tryptophan. *Curr. Opin. Struct. Biol.* **9**, 428-432.

Grosset, J. (1996). Current problems with tuberculosis treatment. *Res. Microbiol.* **147**, 10-16.

Harp, J. M., Hanson, B. L., Timm, D. E. & Bunick, G. J. (1999). Macromolecular crystal annealing: evaluation of techniques and variables. *Acta Crystallog. sect. D,* **55**, 1329-1334.

Harth, G. & Horwitz, M. A. (1999). 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.* **189**, 1425-1436.

Harth, G., Lee, B. Y., Wang, J., Clemens, D. L. & Horwitz, M. A. (1996). Novel insights into the genetics, biochemistry, and immunocytochemistry of the 30-kilodalton major extracellular protein of *Mycobacterium tuberculosis* [published erratum appears in Infect. Immun. (1997), **65**(2), 852]. *Infect. Immun.* **64**, 3038-3047.

Harth, G., Lee, B. Y. & Horwitz, M. A. (1997). High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major *Mycobacterium tuberculosis* extracellular proteins considered to be leading vaccine candidates and drug targets. *Infect. Immun.* **65**, 2321-2328.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature* **381**, 272.

Horwitz, M. A., Lee, B. W., Dillon, B. J. & Harth, G. (1995). Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **92**, 1530-1534.

Horwitz, M. A., Harth, G., Dillon, B. J. & Masleša-Galić, S. (2000a). New recombinant vaccines against tuberculosis more potent than BCG vaccine. *Abstracts of the international meeting: Novel Approaches for Immunization Against Infectious Agents, sponsored by the Israel Science Foundation, Jerusalem, Israel, September 17-19, 2000*, p. 11.

Horwitz, M. A., Harth, G., Dillon, B. J. & Masleša-Galić, S. (2000b). Recombinant BCG vaccines expressing the *Mycobacterium tuberculosis* 30 kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model. *Proc. Natl Acad. Sci. USA* **97**, 13853-13858.

Jackson, M., Raynaud, C., Lanéelle, M. A., Guilhot, C., Laurent-Winter, C., Ensergueix, D., Gicquel, B. & Daffé, M. (1999). Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. *Mol. Microbiol.* **31**, 1573-1587.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallog. sect. A,* **32**, 922-923.

Kleywegt, G. & Jones, T. A. (1998). Databases in protein crystallography (CCP4 Proceedings). *Acta Crystallog. sect. D,* **54**, 1119-1131.

Kraulis, P. J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Kucers, A. & Bennett, N. M. (1987). *The Use of Antibiotics*, 4th edit., J.B. Lippincott Co., Philadelphia, PA.

Lee, B.-Y. & Horwitz, M. A. (1995). Identification of macrophage and stress induced proteins of *Mycobacterium tuberculosis*. *J. Clin. Invest.* **96**, 245-249.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

McRee, D. (1993). *Practical Protein Crystallography*, Academic Press, San Diego, CA.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D,* **53**, 240-255.

Naito, M., Ohara, N., Matsumoto, S. & Yamada, T. (1998). The novel fibronectin-binding motif and key residues of mycobacteria. *J. Biol. Chem.* **273**, 2905-2909.

Nardini, M. & Dijkstra, B. W. (1999). Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr. Opin. Struct. Biol.* **9**, 732-737.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

Otwinowski, Z. & Minor, W. (1997). Processing of diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

Parkin, S. & Hope, H. (1998). Macromolecular cryocrystallography: cooling, mounting, storage, and transporation of crystals. *J. Appl. Crystallog.* **31**, 945-953.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain conformations. *J. Mol. Biol.* **7**, 95-99.

Ronning, D. R., Klabunde, T., Besra, G. S., Vissa, V. D., Belisle, J. T. & Sacchettini, J. C. (2000). Crystal structure of the secreted form of antigen 85C reveals

potential targets for mycobacterial drugs and vaccines. *Nature Struct. Biol.* **7**, 141-146.
Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high resolution refinement. *Methods Enzymol.* **277**, 319-343.
Souhassou, M., Lecomte, C., Blessing, R. H., Aubry, A., Rohmer, M.-M., Wiest, R., Benard, M. & Marraud, M. (1991). Electron distributions in peptides and related molecules. 1. An experimental and theoretical study of N-acetyl-L-tryptophan methylamide. *Acta Crystallog. sect. B,* **47**, 253-266.
Tickle, I. J., Laskowski, R. A. & Moss, D. S. (1998). R_{free} and the R_{free} ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement. *Acta Crystallog. sect. D,* **54**, 547-557.

Edited by I. A. Wilson

(Received 21 August 2000; received in revised form 5 January 2001; accepted 5 January 2001)