nature
structural &
molecular biology

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

# Identification, function and structure of the mycobacterial sulfotransferase that initiates sulfolipid-1 biosynthesis

Joseph D Mougous[1,4], Christopher J Petzold[2], Ryan H Senaratne[3], Dong H Lee[2,4], David L Akey[1], Fiona L Lin[2,4], Sarah E Munchel[1], Matthew R Pratt[2,4], Lee W Riley[3], Julie A Leary[2], James M Berger[1] & Carolyn R Bertozzi[1,2,4]

**Sulfolipid-1 (SL-1) is an abundant sulfated glycolipid and potential virulence factor found in *Mycobacterium tuberculosis*. SL-1 consists of a trehalose-2-sulfate (T2S) disaccharide elaborated with four lipids. We identified and characterized a conserved mycobacterial sulfotransferase, Stf0, which generates the T2S moiety of SL-1. Biochemical studies demonstrated that the enzyme requires unmodified trehalose as substrate and is sensitive to small structural perturbations of the disaccharide. Disruption of *stf0* in *Mycobacterium smegmatis* and *M. tuberculosis* resulted in the loss of T2S and SL-1 formation, respectively. The structure of Stf0 at a resolution of 2.6 Å reveals the molecular basis of trehalose recognition and a unique dimer configuration that encloses the substrate into a bipartite active site. These data provide strong evidence that Stf0 carries out the first committed step in the biosynthesis of SL-1 and establish a system for probing the role of SL-1 in *M. tuberculosis* infection.**

Sulfation of biomolecules is a modification that regulates extracellular interactions by altering the binding properties of the underlying macromolecule[1,2]. Sulfation has been compared to phosphorylation, which by contrast regulates processes that occur inside cells. Sulfotransferases (STs) are the enzymes responsible for transferring the sulfuryl group from a universal donor molecule, 3′-phosphoadenosine-5′-phosphosulfate (PAPS), to carbohydrates, proteins and a variety of other low-molecular-mass metabolites[3–5]. Just as kinases are of considerable interest as mediators of intracellular signaling, sulfotransferases have attracted attention as mediators of intercellular signaling.

Studies of this enzyme class have been most extensively detailed in eukaryotes. Eukaryotic STs are divided into two categories based on their subcellular localization and physiological substrates. Cytosolic STs add sulfate to steroids and other small hydrophobic metabolites, whereas Golgi-resident STs modify carbohydrates and tyrosine residues on membrane-bound and secreted proteins. Examples of processes regulated by mammalian sulfotransferases include interaction of the leukocyte adhesion molecule L-selectin with its endothelial ligands[6,7], binding of HIV gp120 to the coreceptor CCR5 (ref. 8) and to neutralizing antibodies[9], and clearance of glycoprotein hormones[10].

Relatively little information has emerged regarding the pervasiveness and functions of STs and their sulfated products in bacteria[1]. Only two bacterial STs, NodH[11] and NoeE[12] from *Sinorhizobium meliloti*, have been assigned a specific function. Notably, the roles of these prokaryotic carbohydrate STs and eukaryotic carbohydrate STs are analogous. NodH and NoeE direct extracellular interactions between the symbiotic bacterium and its plant host by sulfating the carbohydrate portion of secreted root nodulation factors. The plant then recognizes these metabolites in a sulfate-dependent manner; mutants lacking ST activity show altered host specificity. PAPS biosynthesis enzymes have also recently been identified as virulence determinants in the plant pathogen *Xanthomonas oryzae*, suggesting a role for sulfated metabolites in plant-pathogen interactions[13].

Among human pathogenic bacteria, *Mycobacterium tuberculosis* is unusual in its abundant production of sulfated metabolites[14]. Several other mycobacterial species, including *M. avium*, *M. smegmatis* and *M. fortuitum*, have also been shown to synthesize one or more sulfated compounds[14–17]. Consistent with these biochemical data, we have recently identified the *stf* family of sulfotransferases in mycobacterial genomes[1].

SL-1 is the most notable sulfated metabolite from *M. tuberculosis*. Early studies found a correlation between the abundance of SL-1 and the virulence of several strains[18]. Comprising a T2S core modified with four fatty acyl groups (**Fig. 1a**), SL-1 accounts for ~1% of the dry weight of *M. tuberculosis*[19] and its location in the outer envelope has prompted speculation that it may be a virulence factor involved in host-pathogen interactions[20]. *In vitro* assays have shown that SL-1 causes superoxide release in activated human neutrophils and monocytes, and at low SL-1 concentrations the cells were primed for secondary challenge with mitogens[21–23]. Furthermore, SL-1 has been shown to be immunogenic in human patients and has utility as a serodiagnostic marker[24,25]. A biosynthetic precursor of SL-1, termed $SL_{1278}$ based on its measured mass[14,26], has recently been identified as a potent T-cell antigen presented by the nonclassical MHC molecule CD1b[27]. This compound and fully elaborated SL-1 may combine to serve a function in immunomodulation during *M. tuberculosis* infection.

The biosynthetic pathway for SL-1 has not yet been fully elucidated. Genetic approaches have identified candidates for the polyketide

[1]Department of Molecular and Cell Biology, [2]Department of Chemistry, [3]School of Public Health and [4]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA. Correspondence should be addressed to C.R.B (crb@berkeley.edu).

Published online 18 July 2004; doi:10.1038/nsmb802

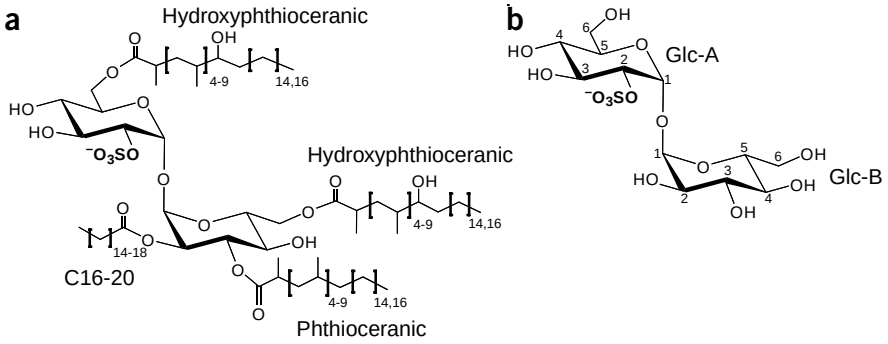


**Figure 1** Structure of sulfolipid-1 and trehalose-2-sulfate. (**a**) The structure of SL-1 as proposed by Goren *et al*[69]. The T2S core of SL-1 is modified with two hydroxyphthioceranic acid groups, a phthioceranic acid group and a palmitic acid group. (**b**) The structure of T2S with residue and atom labeling used in this paper. Before sulfation, trehalose is a $C_2$-symmetric molecule consisting of two glucose moieties in a 1,1-α,α-linkage. The sulfated glucose residue is referred to as Glc-A and the other glucose residue is Glc-B.

synthase that generates the hydroxyphthioceranic acid groups (Pks2, **Fig. 1a**)[28] and a lipid transporter thought to shuttle a precursor of SL-1 across the plasma membrane (MmpL8)[26,29]. In animal models of *M. tuberculosis* infection, mutants lacking MmpL8 showed attenuated virulence, whereas mutants lacking Pks2 behaved similarly to wild-type[26,30]. The presence of SL-1 precursors in these mutant strains complicates the interpretation of the importance of SL-1 in *M. tuberculosis* virulence.

Here we report the characterization of a novel mycobacterial ST. The enzyme, which we termed Stf0 (ref. 1), was identified as a potential ST by virtue of its homology to an ST recently discovered in *S. meliloti*[31]. Biochemical analyses demonstrated that Stf0 transfers the sulfuryl group from PAPS to trehalose to form T2S (**Fig. 1b**), the core disaccharide of SL-1. Kinetic studies with a panel of substrate analogs revealed that the enzyme is highly sensitive to subtle changes in carbohydrate structure. Inactivation of *stf0* in *M. smegmatis* and *M. tuberculosis* confirmed that T2S is the physiological product of Stf0 and is required for the biosynthesis of SL-1. We also determined the crystal structure of *M. smegmatis* Stf0 bound to trehalose. The structure revealed the molecular basis of trehalose recognition and a unique dimer configuration that encloses the substrate into a bipartite active site. Together, these findings identify sulfation of trehalose by Stf0 as the initiating step in SL-1 biosynthesis and provide a framework for defining the role of SL-1 in tuberculosis.

## RESULTS

### Identification of Stf0

A BLAST query of the *M. tuberculosis* genome[32] using a recently identified *S. meliloti* ST LpsS[31] returned an open reading frame (ORF) of moderate homology that has conserved ST PAPS–binding motifs[33]. A single apparent ortholog sharing ~75% identity with this gene (ORF Rv0295c) was also identified in the genome sequences of *M. smegmatis* and *M. avium* (BLAST server at The Institute for Genomic Research; http://www.tigr.org). A complete sequence alignment of the Stf0 orthologs and several related STs is provided in **Supplementary Figure 1** online. Stf0 escaped our previous attempt at an exhaustive ST search in these genomes[1] by virtue of its low homology to other members of the Stf family and the eukaryotic ST search query used at the time.

### Biochemical characterization of Stf0

To test whether Stf0 has ST activity, the gene was amplified from *M. tuberculosis* H37Rv, expressed in *Escherichia coli*, and the protein was purified to homogeneity. Using enzymatically prepared PAP$^{35}$S, we tested whether the purified enzyme was capable of sulfating trehalose *in vitro*. Analysis of the reaction mixture by thin-layer chromatography (TLC) revealed a product with an $R_f$ value identical to that of an authentic sample of T2S generated by chemical synthesis (**Fig. 2a**).

The sulfated product of Stf0 was further analyzed by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). PAP$^{32}$S and PAP$^{34}$S were synthesized enzymatically and these stable isotopically enriched sulfuryl donors were used in larger-scale reactions with Stf0 and trehalose. A major ion with $m/z$ = 421.06, corresponding to the exact mass of sulfated trehalose, was observed in the PAP$^{32}$S reaction (**Fig. 2b**). Likewise, an ion with $m/z$ = 423.06 was identified in the PAP$^{34}$S reaction. These ions were only observed in the presence of trehalose. Analysis of the ion with $m/z$ = 421.06 by MS-MS yielded product ions identical to those obtained from an authentic sample of T2S (data not shown)[14]. These data demonstrate that Stf0 catalyzes the formation of T2S from PAPS and trehalose.

To determine the physiological significance of the above observation, we carried out kinetic studies with trehalose and related natural and unnatural carbohydrates. The $K_m$ value of trehalose was determined to be 18 mM for *M. smegmatis* Stf0 (**Table 1**). This is reasonable for a physiological substrate given that trehalose is present at high concentrations, composing 1–3% of the dry weight of *M. smegmatis*[34]. Modifications to the structure of trehalose severely affected the efficiency of sulfation by Stf0 (**Table 2**). For example, a synthetic analog epimerized at a single stereocenter (Gal(1,1-α,α)Glu) was sulfated 68-fold less efficiently than trehalose. Replacement of one glucose residue of trehalose with a phenyl group, with retention of the α-linkage (α-phenyl glucoside), reduced substrate activity by 3,000-fold versus trehalose. Stf0 showed no detectable activity on the β-linked isomer (β-phenyl glucoside). Thus, Stf0 exhibits a strong preference for an α-glucoside. Free glucose and unnatural stereoisomers of trehalose (α,β (neo-trehalose) and β,β (iso-trehalose)) were not sulfated at a detectable level by Stf0.

**Table 1 Kinetic analysis of *M. smegmatis* Stf0 mutants**

| Enzyme | $K_m$ (mM)[a] | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($s\ M)^{-1}$ |
|---|---|---|---|
| Wild type | 18 ± 2 | 1.6 ± 0.1 | 89 |
| E33A | 29 ± 3 | 0.61 ± 0.02 | 21 |
| E36Q | 6.4 ± 1 | 0.017 ± 0.001 | 2.6 |

[a]All parameters were extracted from a Michaelis-Menten fit ($R^2$ > 0.99) to a minimum of eight trehalose concentrations.

### Stf0 is required for T2S synthesis

Using a combined mass spectrometric and genetic method, we have previously identified T2S as a metabolite in *M. smegmatis*[14]. *Stf0* was disrupted in *M. smegmatis* (Δ*s-stf0*) to test the function of the gene *in vivo*. To probe for changes in sulfated metabolite production, wild-type *M. smegmatis*, Δ*s-stf0*, and a complemented mutant overexpressing Stf0 (Δ*s-stf0:Gstf*) were grown in the presence of $^{35}SO_4^{2-}$. Organic extracts from these cultures were prepared and separated by TLC. A

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

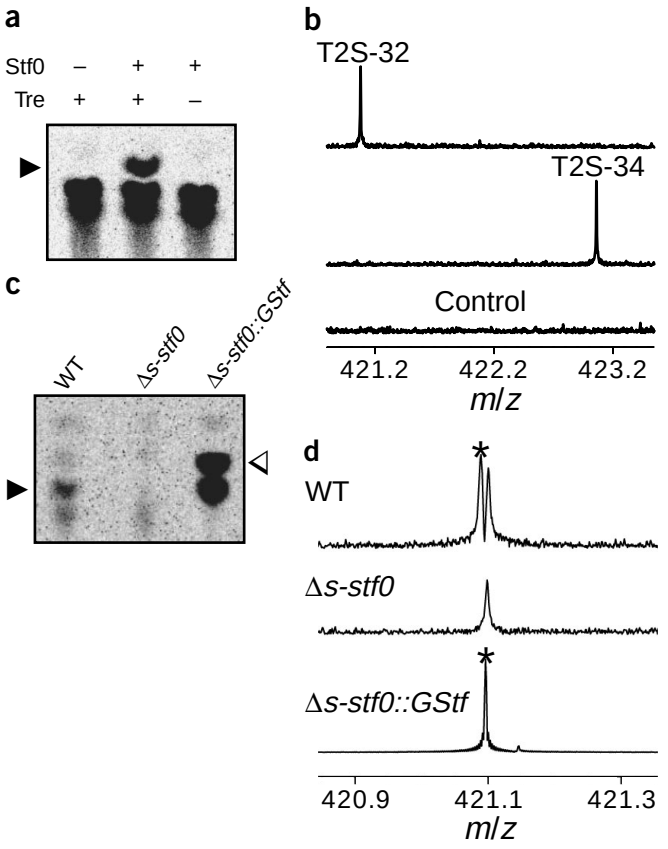


**Figure 2** Stf0 is the mycobacterial T2S sulfotransferase. (**a**) TLC analysis of reactions with purified Stf0, trehalose (Tre) and PAP$^{35}$S. Arrowhead, a sulfated product formed in the experimental lane. (**b**) FT-ICR mass spectra of reactions with purified Stf0, trehalose, and either PAP$^{32}$S (top and bottom spectra) or PAP$^{34}$S (middle spectrum). In the reaction analyzed in the top spectrum, T2S of the $^{32}$S form (T2S-32) is synthesized ($m/z$ = 421.06) and in the reaction analyzed in the middle spectrum, the $^{34}$S form (T2S-34) is made ($m/z$ = 423.07). The control reaction analyzed in the bottom spectrum lacks trehalose. (**c**) TLC analysis of crude extracts from $^{35}SO_4^{2-}$-labeled *M. smegmatis* cultures. Black arrowhead, position of T2S observed in extracts of the wild type (WT) and complemented Stf0 knockout (Δ*s-stf0:Gstf*) strains. The spot corresponding to T2S is absent from extracts of Δ*s-stf0*. White arrowhead, position of additional sulfated molecules that markedly increased in abundance upon overexpression of Stf0. (**d**) FT-ICR mass spectra of crude extracts from the same *M. smegmatis* strains as in **c** showing the region surrounding T2S. The peak corresponding to T2S is highlighted with an asterisk. The neighboring isobar observed in the top two spectra has the exact mass of a phosphorylated disaccharide. In the bottom spectrum, this isobar is masked by the marked overproduction of T2S.

compound with an $R_f$ value matching T2S was found in wild type, but was not visualized in Δ*s-stf0* (**Fig. 2c**). Notably, overexpression of Stf0 in Δ*s-stf0:Gstf* led to a marked increase in levels of T2S as well as other potentially related sulfated compounds. Crude extracts from these cultures were also prepared for MS analysis. FT-ICR MS was chosen, as it can resolve the cellular pool of phosphorylated and sulfated disaccharides[35]. Wild-type extracts possessed a compound with $m/z$ = 421.06, corresponding to T2S, that was absent in Δ*s-stf0* and present in great abundance in Δ*s-stf0:Gst* (**Fig. 2d**). Broad molecular mass scans showed no other gross differences in metabolite profile between the Δ*s-stf0* and wild-type strains (data not shown).

### Stf0 is required for SL-1 biosynthesis

To determine whether Stf0 is responsible for initiating the biosynthesis of SL-1, we disrupted *stf0* in *M. tuberculosis* (Δ*t-stf0*). A crude extract from Δ*t-stf0* was analyzed for the presence of T2S and SL-1 by MS as described[14]. T2S (data not shown) and SL-1 (**Fig. 3a**) were readily observed in wild type, whereas both compounds were absent in Δ*t-stf0*. These data confirm that sulfation of trehalose by Stf0 is required for SL-1 biosynthesis, however MS could not address whether trehalose was elaborated with (hydroxy)phthioceranic acids in the absence of sulfation by Stf0.

To probe for unsulfated SL-1 intermediates, we used $^{14}$C-propionic acid as a metabolic label specific for polypropionates such as (hydroxy)phthioceranic acids[26]. As expected based on our MS results, $^{14}$C-propionic acid labeling of Δ*t-stf0* confirmed the absence of fully elaborated SL-1 (**Fig. 3b**). Moreover, no compounds corresponding to an unsulfated form of either SL-1 or an SL-1 precursor (the presence of the sulfate would affect the $R_f$ of the molecules) were observed. These observations strongly suggest that sulfation of trehalose is an obligate first step in the biosynthesis of SL-1.

**Table 2 Kinetic analysis of *M. smegmatis* Stf0 with trehalose and analogs**

| Substrate | $k_{cat}/K_m$[a] (s M)$^{-1}$ | $(k_{cat}/K_m) / (k_{cat}/K_m)^{tre}$ |
|---|---|---|
| Trehalose | 89 ± 6 | 1.0 |
| Gal(1,1-α,α)Glu | 1.3 ± 0.1 | 0.015 |
| α-Phenyl glucoside | 0.030 ± 0.001 | $3.4 \times 10^{-4}$ |
| β-Phenyl glucoside | <0.01 | ND[b] |
| Glucose | <0.01 | ND |
| α,β-Trehalose | <0.01 | ND |
| β,β-Trehalose | <0.01 | ND |

[a]Parameter was determined using a minimum of four substrate concentrations. Values ± s.e.m. were determined from triplicate measurements. [b]ND, not determined.

### Structure of Stf0 bound to trehalose

To better understand the molecular interactions governing the initiating step of SL-1 biosynthesis, we undertook the X-ray crystal structure of Stf0. Although we were unable to grow crystals of the *M. tuberculosis* ortholog, trehalose-dependent crystals of the *M. smegmatis* enzyme were readily obtained. These crystals belonged to the $P3_221$ space group and contained four monomers in the asymmetric unit (**Supplementary Table 1** online). The four protomers were configured as two identical dimers (**Fig. 4**). Electron density clearly indicated that each protomer bound one trehalose molecule (**Fig. 5a**). The Stf0 monomer belongs to the ST structure superfamily[36] and consists of a single domain with a core four-stranded parallel β-sheet flanked by α-helices (**Fig. 4a**).

ST structures solved prior to this work include several eukaryotic cytosolic STs (such as estrogen ST (EST)), the ST domain of the Golgi-resident heparan sulfate *N*-deacetylase/*N*-sulfotransferase 1 (HDNST-1)[36], and recently, a Golgi-resident heparan sulfate 3-*O*-sulfotransferase[37]. Stf0 is the first prokaryotic ST to be structurally characterized and the first carbohydrate ST to have its structure solved in the presence of an acceptor substrate. Among ST structures, Stf0 shares the most structural homology with HDNST-1 (ref. 38), with an r.m.s. deviation of 3.5 Å across 183 topologically conserved Cα atoms. These conserved residues are found primarily within the core β-sheet, the 5′-phosphosulfate-binding loop (5′ PSB)[33], and helices α7 and α9 (**Fig. 4a**). Based on superimpositions of Stf0 with human HDNST-1 and murine EST[39], helices α7 and α9 are expected to form the majority of interactions with the 3′ phosphate and adenine base of PAPS.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

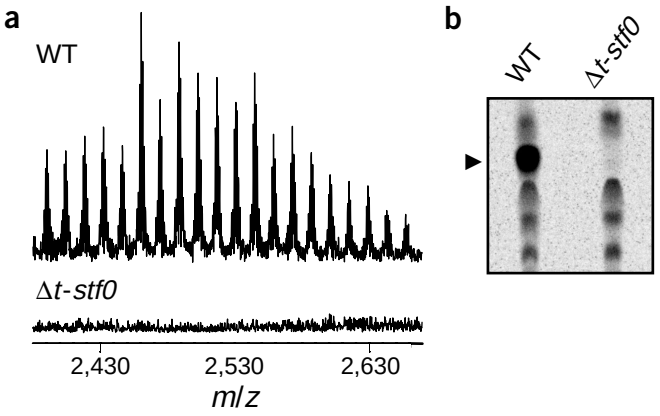


**Figure 3** Sulfation of trehalose is required to initiate SL-1 biosynthesis. (**a**) The SL-1 region of FT-ICR mass spectra from crude extracts of *M. tuberculosis* cultures. SL-1 is characteristically observed as an envelope of lipoforms (±14 mass units) in WT. Under identical conditions, SL-1 is absent from Δ*t-stf0*. (**b**) TLC analysis of crude extracts from $^{14}$C-propionic acid-labeled *M. tuberculosis* cultures. Black arrowhead, position of SL-1. SL-1 is not observed in Δ*t-stf0*, nor are unique spots observed that could correspond to unsulfated SL-1.

A perfectly conserved serine residue is found on α7 of Stf0 (residue 152) and superposes with the equivalent serine residues of human HDNST-1 (Ser712) and murine EST (Ser138). In these ST structures, the conserved serine forms a hydrogen bond with an oxygen atom of the 3′ phosphate of PAPS.

### The Stf0 dimer

Many eukaryotic cytosolic STs have been shown to exist as homo- and heterodimers in solution. A recent study has demonstrated that the dimer interface in these proteins is restricted to a conserved ten-residue loop located near the C terminus[40]. To assess the oligomeric state of Stf0, we prepared native (His-tag-free) samples of the protein for analytical equilibrium centrifugation studies. Results of these experiments indicated that Stf0 exists predominantly as a homodimer in solution (**Supplementary Fig. 2** online). To our knowledge, Stf0 is the first carbohydrate ST that has been shown to oligomerize.

A dimer of Stf0 protomers was clearly visible in the structure (**Fig. 4c**). Notably, the protomer contacts and the overall dimer configuration observed in the structure of Stf0 bear no similarity to those of the eukaryotic cytosolic ST dimers. Nonetheless, several lines of evidence give us confidence that the observed Stf0 dimer is physiologically relevant. The trehalose-bound $P3_221$ crystal form presented here revealed four monomers in the crystallographic asymmetric unit. Only one dimer interface, that assigned here, was shared by all crystallographically and noncrystallographically related protomers. In addition, we have solved a second crystal form of Stf0 in the absence of trehalose that crystallizes into an alternative space group bearing the Stf0 dimers we observed here (data not shown).

Stf0 dimer contacts bury 1,370 Å$^2$ or 11.5% of the solvent-accessible surface area of each monomer. The interactions are largely symmetrical and encompass the trehalose-binding site of both protomers (**Fig. 4c**). A large proportion of the Stf0 dimer contacts are polar (**Fig. 4b**). The region of the dimer interface located directly

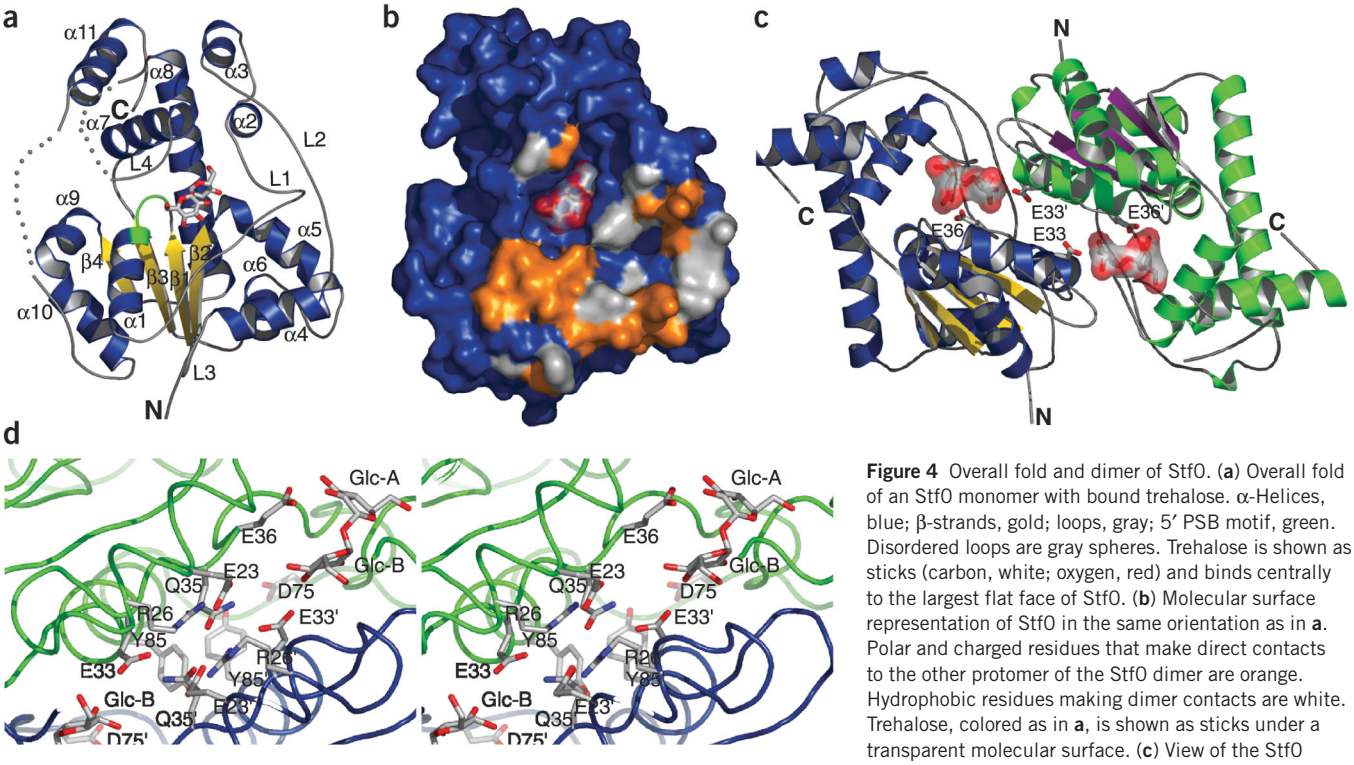


**Figure 4** Overall fold and dimer of Stf0. (**a**) Overall fold of an Stf0 monomer with bound trehalose. α-Helices, blue; β-strands, gold; loops, gray; 5′ PSB motif, green. Disordered loops are gray spheres. Trehalose is shown as sticks (carbon, white; oxygen, red) and binds centrally to the largest flat face of Stf0. (**b**) Molecular surface representation of Stf0 in the same orientation as in **a**. Polar and charged residues that make direct contacts to the other protomer of the Stf0 dimer are orange. Hydrophobic residues making dimer contacts are white. Trehalose, colored as in **a**, is shown as sticks under a transparent molecular surface. (**c**) View of the Stf0 dimer looking down the two-fold axis. One protomer is colored as in **a**. The other protomer is colored with green α-helices, purple β-strands and gray loops. Trehalose is represented in stick form under a partially transparent molecular surface. Two amino acids from each protein are represented in stick form, Glu33 and Glu36. Glu36 is the proposed active site base and resides on the same loop as Glu33. Glu33 reaches into the active site of the other protomer where it hydrogen-bonds to the 4-OH group of Glc-B. (**d**) Stereo diagram of the Stf0 dimer interface. The two protomers are blue and green tubes. Residues involved in dimer interactions are in stick model.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

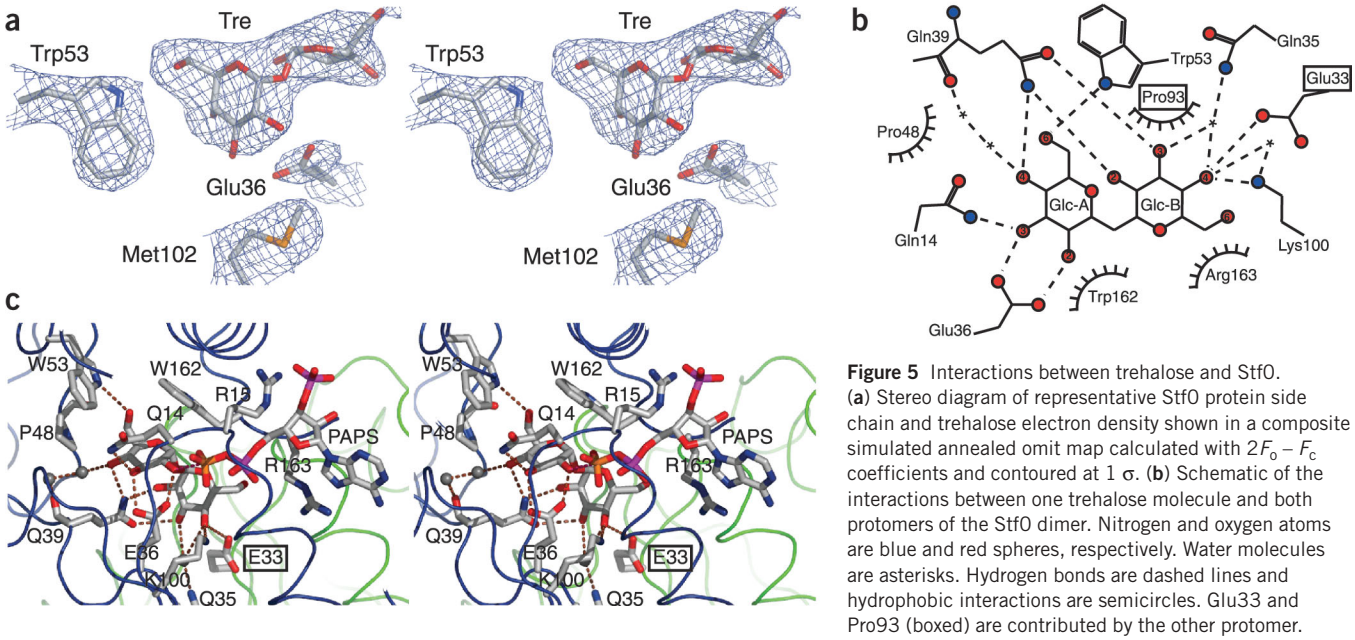


**Figure 5** Interactions between trehalose and Stf0. (**a**) Stereo diagram of representative Stf0 protein side chain and trehalose electron density shown in a composite simulated annealed omit map calculated with $2F_o - F_c$ coefficients and contoured at 1 σ. (**b**) Schematic of the interactions between one trehalose molecule and both protomers of the Stf0 dimer. Nitrogen and oxygen atoms are blue and red spheres, respectively. Water molecules are asterisks. Hydrogen bonds are dashed lines and hydrophobic interactions are semicircles. Glu33 and Pro93 (boxed) are contributed by the other protomer. For a table containing the lengths of all hydrogen bonds shown in this figure or discussed in the text, see **Supplementary Table 2** online. (**c**) Stereo view of the Stf0 active site with modeled PAPS. Cα traces of the two protomers are blue and green tubes. Aside from the addition of the 5′ PSB residue Arg15 and the removal of Pro93, residues represented as sticks are the same as those detailed in **a**. Hydrogen bonds are dashed lines and water molecules are gray spheres. PAPS was modeled based solely on a superposition of an EST–PAPS complex and Stf0. A pink dashed line shows the close proximity of the PAPS sulfur atom (orange) to the 2-OH group of Glc-A.

between the trehalose molecules is particularly hydrophilic (**Fig. 4d**). Key residues involved in this extensive electrostatic and hydrogen-bonding network are located on α1 (Glu23 and Arg26) and L1 (Glu33 and Gln35). A molecular surface representation of the Stf0 dimer suggests that the entry of trehalose and exit of T2S would be impeded by unfavorable steric interactions with residues at the dimer interface. Studies are underway to determine how trehalose gains access to the active site. The hydrophilic nature of the dimer interface suggests that a substantial percentage of the protein may be in the monomeric form, whereupon trehalose could bind and T2S could exit the protein without major structural changes.

Tyr85 is an exception to the high overall degree of symmetry found at the Stf0 dimer interface (**Fig. 4d**). The residue adopts alternative rotomers in each protomer, facilitating a Tyr85-Tyr85′ (prime denotes the second protomer of the dimer) π-stacking interaction. This disruption of local symmetry leads to the inability of one of the tyrosine residues to take part in a second dimer contact, a hydrogen bond with Asp75′ located on L2 (**Fig. 4d**). Local asymmetry at symmetric dimer interfaces has recently been reviewed[41] and can indicate cooperativity, allosteric regulation, or more commonly, optimized packing constraints near a symmetry axis. Tyr85 is located on the two-fold symmetry axis and makes no obvious contacts with active site residues in either conformation, suggesting that optimized packing constraints drive the asymmetry.

Glc-B of trehalose (defined in **Fig. 1b** and in the following section) also contributes part of the dimer interface. Glu33 on L1 of one protomer extends into the active site of the other, where it forms hydrogen bonds with Glc-B (**Figs. 4c** and **5b,c**). The distance of this interaction varies substantially in its four occurrences in the asymmetric unit of the structure. In one instance, the residue is in close enough proximity to make strong hydrogen bonds with both the 4-OH and 3-OH groups of Glc-B, whereas in two other protomers, Glu33 is within hydrogen-bonding distance only to the 4-OH group. In the final case, Glu33 is

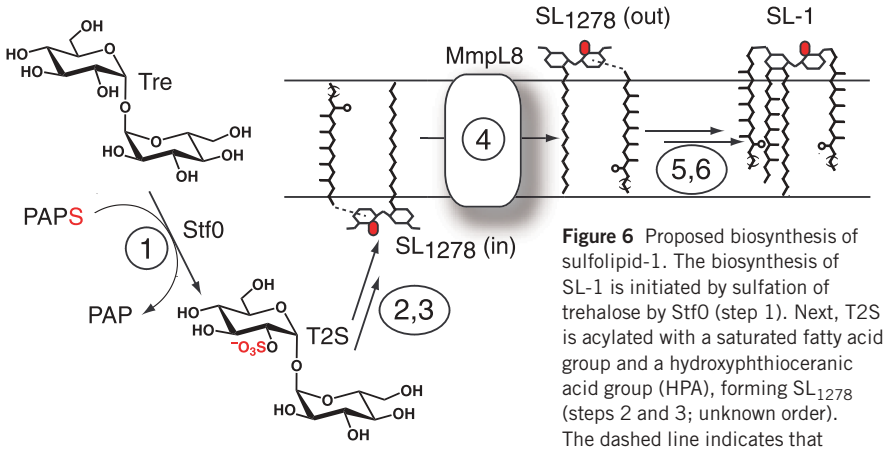


**Figure 6** Proposed biosynthesis of sulfolipid-1. The biosynthesis of SL-1 is initiated by sulfation of trehalose by Stf0 (step 1). Next, T2S is acylated with a saturated fatty acid group and a hydroxyphthioceranic acid group (HPA), forming $SL_{1278}$ (steps 2 and 3; unknown order). The dashed line indicates that the position of HPA on $SL_{1278}$ is uncertain[26,27,29]. Gilleron *et al.*[27] propose a structure in which HPA occurs at the 3′-OH of Glc-B whereas the structure of SL-1 proposed by Goren *et al.*[69] implies that the HPA group of $SL_{1278}$ occurs on the 6- or 6′-position. $SL_{1278}$ is next transported across the plasma membrane by MmpL8 (step 4). After transport by MmpL8, the biosynthesis of SL-1 is completed by acylation of $SL_{1278}$ with a phthioceranic acid group and another HPA group (steps 5 and 6; unknown order). Aspects of the pathway are reviewed in ref. 70.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

just outside of predicted hydrogen-bonding distance to the 4-OH group of Glc-B. Notably, our biochemical data with the trehalose analog Gal(1,1-α,α)Glu support a role for the 4-OH of Glc-B as a crucial determinant of substrate recognition (**Table 2**).

We also introduced a E33A mutation and measured the kinetic parameters with respect to trehalose (**Table 1**). The $K_m$ value increased, suggesting that interactions with trehalose had been compromised. A more marked effect was observed with $k_{cat}$ (reduced three-fold), suggesting that this residue affects catalytic efficiency despite its position distal from the site of chemical transformation. Given its location at the dimer interface and placement on the same loop as Glu36 (L1), which we propose serves as the active site base (see below), Glu33 may serve a role in coordinating dimer formation with substrate binding and catalysis. L1 replaces the second core β-strand found in most STs and provides for many unique features of Stf0. L1 not only links the two protomer active sites through contacts with both trehalose molecules (via Glu33 and Glu36), but also provides contacts via Gln35 to a crucial water molecule that coordinates the 3- and 4-OH groups of Glc-B (**Fig. 5b,c**).

### Stf0-trehalose interactions

Trehalose is a symmetric molecule that is desymmetrized by sulfation. It is therefore important to distinguish between the glucose moiety that is to be sulfated (Glc-A) and the other glucose, which remains unaltered (Glc-B; **Fig. 1b**). Based on proximity to the conserved PAPS-binding site, and on specific contacts made to the nucleophilic 2-OH group (see below), Glc-A of trehalose can be unambiguously identified.

The closest interactions between trehalose and Stf0 are a pair of hydrogen bonds between the side chain oxygen atoms of Glu36 and the 2- and 3-OH groups of Glc-A (**Figs. 4c,d** and **5b,c**). Because no other residues of Stf0 contact the acceptor hydroxyl of trehalose (2-OH), we postulate that Glu36 is likely to serve as the catalytic base for an $S_N2$-like in-line displacement mechanism that has been widely reported for other STs[36,42]. A E36Q mutation reduced $K_m$ by 3-fold and $k_{cat}$ by ~100-fold (**Table 1**). The marked effect on $k_{cat}$ is consistent with a role in the chemical step, although we were surprised to observe substantial residual activity given previous reports that comparable mutants of other STs are inactive[37,43]. It is possible that the Stf0-catalyzed reaction is more dissociative and therefore less sensitive to modulation of its active site base. Few studies have addressed the mechanistic details of STs, but one recent report assigned a dissociative mechanism to β-arylsulfotransferase IV[44].

Superposition of Stf0 with the structure of EST solved in the presence of both PAPS and the sulfate acceptor estradiol (E2)[45] provides further support for the proposed in-line sulfuryl transfer mechanism. The PAPS molecule modeled directly from the superposition has minimal unfavorable steric interactions with Stf0 and crucial residues in the 5′ PSB motifs are well positioned to bind PAPS in the expected manner (**Fig. 5c**). The sulfur atom of PAPS is 3 Å from the 2-OH group of Glc-A; this is highly indicative of a direct transfer mechanism. This superposition also led to a near-perfect overlap of the acceptor hydroxyl groups of E2 and T2S (data not shown). The marked proximity of these groups (0.6 Å) is noteworthy given the chemical diversity of trehalose and E2, and the evolutionary distance between Stf0 and EST.

Each hydroxyl group of Glc-A is recognized by the enzyme through at least one hydrogen bond to either a well-ordered water molecule or a side chain (**Fig. 5b,c**). Gln39 forms particularly extensive contacts with both glucose moieties of trehalose. As is commonly observed in carbohydrate-binding pockets[46], an aromatic residue (Trp162) forms the lid of the active site through hydrophobic interactions with the relatively apolar face of Glc-A. Pro48 defines the backside of the active site by contacting C5 of Glc-A (**Fig. 5b,c**). The region of the active site occupied by Glc-B is composed of residues from both protomers in the dimer (**Fig. 5b,c**). The surface of this distal portion of the active site is defined primarily by Arg163 and Pro93′. Additionally, Glu33′ (discussed above), Lys100, and two ordered water molecules take part in a hydrogen-bonding network to Glc-B. The extensive interactions of the trehalose hydroxyl groups with amino acid side chains or bound water molecules suggest that any modifications, such as acylation, would preclude enzyme binding. This observation has implications for the biosynthetic order of the modifications found in SL-1, as discussed below.

## DISCUSSION

The biochemical, genetic and structural data presented here identify Stf0 as the first committed step in the biosynthesis of SL-1. Stf0 shows a strong preference for trehalose and minor structural alterations to trehalose have a marked effect on catalytic efficiency. Disruption of *stf0* in two mycobacterial species led to a loss of T2S, as observed by mass spectrometry and TLC analysis. SL-1 was no longer detected in the *M. tuberculosis* mutant, nor were unsulfated acylated precursors. Finally, the structure of Stf0 revealed a bound trehalose molecule that was fully enclosed at a unique dimer interface. The mode of binding observed in the structure precludes elaboration of the trehalose hydroxy groups; this is consistent with sulfation as the first step in SL-1 biosynthesis. Based on our current work and other recent studies, we propose a biosynthetic route for SL-1 (**Fig. 6**). Although three mutants in SL-1 biosynthesis have now been constructed, many details remain unresolved. For example, the order of lipid addition to form $SL_{1278}$, the mechanism and extent of $SL_{1278}$ translocation by MmpL8, and the order of addition of the final two lipids are not known.

The correlation of SL-1 abundance to virulence in field isolates, its elaborate chemical structure, and its restricted distribution to *M. tuberculosis* have motivated many efforts to elucidate its function. Thus far, two independent studies with Δpks2 (refs. 26,30), a mutant unable to synthesize the hydroxyphthioceranic acids of SL-1 (**Fig. 6**), have shown that the fully elaborated molecule is not required for growth in both the mouse and guinea pig. One possible explanation for this result is that SL-1 precursors, such as palmitoylated T2S, may accumulate in the mutant and may be sufficient to carry out the function of the entire molecule. Indeed, some SL-1 precursors, notably $SL_{1278}$ and a synthetic analog (2,3-dipalmitoyl-trehalose-2-sulfate), are potent antigens capable of eliciting interferon-γ production in CD1b-restricted T-cell clones from tuberculosis patients[27]. Another explanation is that SL-1 may be highly evolved to target features of the human host that are not present in the mouse. The SL-1 knockout strain we describe here will be an important tool for addressing these questions.

The finding that sulfation of trehalose affects further SL-1-specific elaboration indicates a function for this modification in directing the compound into alternative biosynthetic pathways. This may be a means to accommodate the two distinct roles that trehalose plays in mycobacteria, one in free form as an osmoprotectant and protein stabilizer, and another as a component of complex molecules such as SL-1 and other cell-wall glycolipids[47]. To function as an osmoprotectant and protein stabilizer, trehalose must be maintained at high levels and undergo substantial fluctuations depending on the status of the cell[48]. To remain in homeostasis under these conditions, some trehalose-based metabolites may require an additional level of control. In this model, sulfation acts in a manner analogous to phosphorylation of carbohydrates, which is often used to direct metabolic flux.

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

Stf0 adds to a growing list of structures of bacterial enzymes that modify trehalose[49,50]. The structure of Stf0 reveals extensive contacts made with both monosaccharides of trehalose, thereby explaining how the enzyme distinguishes this substrate from related carbohydrates in the cell. The energetic cost of desolvating trehalose for recognition by the enzyme is overcome by protein-protein interactions that fully encapsulate the substrate at a unique dimer interface. This first structure of a bacterial ST has a notable degree of similarity in active site configuration with distant mammalian STs. Although the sequence, oligomeric configuration, and chemical nature of the substrates of these enzymes are divergent, several critical active site residues and the nucleophilic acceptor hydroxyls are essentially superimposable.

The identification of Stf0 and the X-ray structure establish a platform for defining the physiological functions of SL-1 using chemical and genetic approaches. *M. tuberculosis* deficient in Stf0 lack both SL-1 and partially modified analogs and therefore may not suffer complications associated with *in vivo* studies of the *pks2* and *mmpL8* mutants. Chemical strategies for studying SL-1-dependent processes include the use of small molecule inhibitors of Stf0 to block SL-1 biosynthesis in a temporally controlled fashion during infection. The availability of an SL-1 structure, combined with recently reported library screening approaches[51,52], should facilitate this effort.

## METHODS

**Stf0 knockout construction.** *M. smegmatis*[53] *stf0* was identified in the preliminary genome sequence using Rv0295c (*M. tuberculosis* Stf0) to query the BLAST server available at The Institute for Genomic Research (http://www.tigr.org). The *M. smegmatis* Δ*s-stf0* strain was constructed as described[54]. Genomic DNA upstream of the deleted region of the *stf0* gene was amplified by PCR using the forward primer 5′-AAGCTTGACCCGCCCGAGCGGCCGTTC-3′ and the reverse primer 5′-GGATCCGTACTGGAAGAACTCCTGCGG-3′. The downstream flanking genomic DNA was amplified by PCR using the forward primer 5′-GGATCCCTCGGCCAGGACCCACGCCTG-3′ and the reverse primer 5′-CGGCCGCGATGTGCGCCTTGAAGATCTGC-3′. After digestion with the appropriate restriction enzymes, these PCR products were subcloned into the p2NIL vector. The resulting mutant allele encodes an in-frame deletion of residues 40–228.

*M. tuberculosis* (American Type Culture Collection no. 27294) *stf0* was identified as ORF Rv0295c in the published genome sequence[32] using an *S. meliloti* LPS ST sequence shared by David Keating as a query in the 'Tuberculist' BLAST server (http://genolist.pasteur.fr/TubercuList). The knockout (Δ*t-stf0*) was constructed using the same strategy as used for Δ*s-stf0*. Upstream flanking DNA was amplified by PCR using the forward primer 5′-AAGCTTCGGTTGTGCGATCACAAAGGAC-3′ and the reverse primer 5′- TCTAGAGCTGCGTTGGGTGGCGAGCAC-3′. Downstream flanking DNA was amplified by PCR using the forward primer 5′-TCTAGAGTGGCCAGTGTGCTTGACGCC-3′ and the reverse primer 5′- TTAATTAAGCGCGTGCACAATGTTGTTGGC-3′. The resulting mutant allele encodes an in-frame deletion of residues 19–221.

For complementation of Δ*stf0*, the mutant was electroporated with a plasmid that overexpresses *M. smegmatis stf0* (pMSGS_Stf0). The plasmid is a derivative of pMS2Kan[55] that contains the promoter and the first three amino acids of glutamine synthase (GS)[56]. Genes were cloned between the GS leader peptide (N-terminal) and a FLAG epitope (C-terminal), facilitating robust constitutive expression that could be monitored.

**Metabolic labeling of mycobacterial cultures.** For *M. smegmatis*, 125 μCi of $Na_2{}^{35}SO_4$ (ICN Pharmaceuticals) was added to 5 ml of log phase culture in 7H9 media. The culture was grown for an additional 24 h before being harvested by centrifugation at 2,000*g* and extracted with 1 ml of 2:1 chloroform/methanol as described below. Clarified extract (1 μl) was spotted onto a silica gel TLC plate and separated with 1:3:1 $H_2O/NH_4OH/i$-propanol. *M. tuberculosis* labeling was done by adding 3.3 μCi of $^{14}$C-labeled propionic acid to 5 ml freshly inoculated 7H9 media. The cultures were grown to late log phase before being heat-killed, harvested and extracted as described above for *M. smegmatis*. Labeled lipids were separated with 60:12:1 chloroform/methanol/water on silica gel TLC plates.

**Mass spectrometry.** For FT-ICR MS analysis, a 2 ml culture of each strain was grown to late log phase in 7H9 medium. The cells were then pelleted by centrifugation at 2,000*g* and resuspended in 0.5 ml of 2:1 chloroform/methanol. This slurry was vigorously shaken for 5 h at room temperature. The organic phase was clarified by centrifugation at 20,000*g* and removed for MS analysis.

Mass spectra were acquired on a Bruker Apex II FT-ICRMS (Bruker Daltonics) equipped with a 7 T actively shielded magnet. Cell extracts, in 2:1 chloroform/methanol solution, were introduced into the mass spectrometer at 2 μl $min^{-1}$ by means of an Analytica (Analytica) or Apollo (Bruker Daltonics) pneumatically assisted electrospray source in negative ion mode. After ionization, ions were accumulated in an rf-only external hexapole for 2 s before being transferred to the ICR cell for mass analysis. Spectra are composed of 512 k and 1 M data points and are an average of between 4 and 32 scans. The spectra were acquired on the FT-ICR data station, operating Xmass version 5.0.10 (Bruker Daltonics).

**Protein expression and purification.** The ORFs encoding *M. smegmatis* and *M. tuberculosis* Stf0 were amplified by PCR, cloned into pET28B, and the resulting plasmids were transformed into the *E. coli* strain BL21. Mutants were generated using QuikChange (Stratagene) mutagenesis and the sequences were confirmed by DNA sequencing. Protein expression was carried out with individual transformants overnight at 16 °C with 0.35 mM IPTG. Cells were harvested and resuspended in 15 ml of lysis buffer (50 mM Tris, pH 7.5, 0.5 M NaCl, 10% (v/v) glycerol, 15 mM imidazole, 2 mM b-mercaptoethanol, protease inhibitors) per liter of culture. Stf0 was purified under standard conditions using a PorosMC (Perspective Biosystems) column followed by a concentration step and an S-300 sizing column. $His_6$-tagged *M. smegmatis* Stf0 was concentrated to 20 mg $ml^{-1}$, stored and dialyzed into a minimal solution of 100 mM NaCl, 10 mM Tris, pH 7.5, and 1 mM TCEP for crystallography purposes. Selenomethionine protein was overexpressed in minimal media using a described protocol[57] and purified similarly to native Stf0. Unlike native Stf0, the selenomethionine protein was unstable and would not crystallize after ~3 d from cell lysis. These crystals were small and required microseeding with native protein.

**Crystallography.** Crystals of *M. smegmatis* Stf0 were obtained using vapor diffusion by mixing 1 μl of dialyzed protein (see above), 1 μl of a well solution (13% (v/v) PEG 3K and 100 mM citrate, pH 5.2), and 0.2 μl of 1 M trehalose. The crystals were slowly introduced into a final cryoprotectant solution consisting of well solution, 1 mM TCEP, 75 mM NaCl, and 20% (v/v) PEG 400, by increasing PEG concentration stepwise (2, 5, 10, 15, 20%; 5 min per step).

Data were collected at beamlines 8.2.2 and 8.3.1 at the Advanced Light Source using an ADSC Quantum-Q315 or Q210 CCD detector, respectively. Diffraction data were processed using HKL2000 (ref. 58) (**Supplementary Table 1** online). A two-wavelength data set from a single selenomethionine protein crystal was used for phasing. Selenium positions were initially found in a SAD experiment using SOLVE[59]. These sites were then refined using both the peak and remote wavelengths with MLPHARE[60] running under ELVES[61]. RESOLVE was used to extend the resolution to native data and to calculate initial experimental electron density maps. Model building was carried out with O[62] and refinement with REFMAC5[63] using TLS[64] and NCS restraints. A large fraction (15%) of the residues in the asymmetric unit were disordered and therefore not included in the final model. Water molecules were added with ARP/WARP[65]. Final $R_{work}$ and $R_{free}$ values of 21.8% and 27.5%, respectively, were obtained. All molecular figures were generated with PyMOL (http://www.pymol.org). Calculations for buried surface area were carried out using AREAIMOL[60] with default parameters and a 1.4-Å sphere. Dimer contacts were compiled with CONTACT[60] using an intersubunit distance cutoff of 3.6 Å.

**Sulfotransferase assays.** Stf0 was purified as described above. After purification, the $His_6$-tag was removed by incubating with thrombin (1 U $mg^{-1}$ of protein) for 4 h at room temperature. The reaction was quenched with 1 mM PMSF (Sigma), and the impurities and uncleaved protein were removed by

passing the protein over a PorosMC column, followed by concentration and gel filtration as described above. PAP$^{35}$S was synthesized as described[66] using recombinant purified *E. coli* APS kinase and *S. cerevisiae* ATP sulfurylase (Sigma). PAP$^{34}$S and PAP$^{32}$S were prepared similarly to PAP$^{35}$S, with the addition of either 10 mM $Na_2{}^{34}SO_4$ (ICON Isotopes) or $Na_2{}^{32}SO_4$, respectively. Radioactive ST reactions were done at room temperature in 20 mM Tris, pH 7.5, 50 mM NaCl, 0.5 μCi PAP$^{35}$S, and 100 μM PAPS. All analogs, with the exception of Gal(1,1-α,α)Glu, were commercially available. The synthesis of Gal(1,1-α,α)Glu has been described[67]. Reaction products were separated with silica gel TLC plates using 1:3:1 $H_2O$/2-propanol/$NH_4OH$ as a solvent system. The $R_f$ of the sulfated product of the enzymatic reaction was compared to a chemically synthesized T2S standard[68]. Reaction products were quantified using phosphorimaging followed by densitometry. A 100 μl reaction containing 100 mM ammonium bicarbonate, pH 8.5, 2 mM trehalose, 100 ng Stf0, and 0.6 mM PAP$^{32}$S or PAP$^{34}$S that proceeded for 1 h was characterized by MS.

**Coordinates.** The coordinates and structure factors of Stf0 have been deposited in the Protein Data Bank (accession code 1TEX).

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

ACKNOWLEDGMENTS

We thank the TB Structural Genomics Consortium (http://www.doe-mbi.ucla.edu/TB) for partial funding. We are grateful to J. Holton and C. Ralston at the Advanced Light Source for their assistance with data collection and processing, D. Chen and M. Schelle for fruitful scientific discussions, members of the Berger and Bertozzi labs for careful review of the manuscript, D. Keating for sharing valuable sequence data and H. Schachman for use of equipment. J.D.M. was supported by a fellowship from the Ford Foundation. This work was supported by a grant from the US National Institutes of Health to C.R.B. (AI51622) and J.M.B. (P50-GM62410).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 8 March; accepted 19 May 2004
Published online at http://www.nature.com/nsmb/

1. Mougous, J.D., Green, R.E., Williams, S.J., Brenner, S.E. & Bertozzi, C.R. Sulfotransferases and sulfatases in mycobacteria. *Chem. Biol.* **9**, 767–776 (2002).
2. Hemmerich, S. & Rosen, S.D. Carbohydrate sulfotransferases in lymphocyte homing. *Glycobiology* **10**, 849–856 (2000).
3. Honke, K. & Taniguchi, N. Sulfotransferases and sulfated oligosaccharides. *Med. Res. Rev.* **22**, 637–654 (2002).
4. Moore, K.L. The biology and enzymology of protein tyrosine O-sulfation. *J. Biol. Chem.* **278**, 24243–24246 (2003).
5. Coughtrie, M.W. Sulfation through the looking glass—recent advances in sulfotransferase research for the curious. *Pharmacogenomics J.* **2**, 297–308 (2002).
6. Hemmerich, S., Butcher, E.C. & Rosen, S.D. Sulfation-dependent recognition of high endothelial venules (HEV)-ligands by L-selectin and MECA-79, and adhesion-blocking monoclonal antibody. *J. Exp. Med.* **180**, 2219–2226 (1994).
7. Hemmerich, S. *et al.* Sulfation of L-selectin ligands by an HEV-restricted sulfotransferase regulates lymphocyte homing to lymph nodes. *Immunity* **15**, 237–247 (2001).
8. Farzan, M. *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* **96**, 667–676 (1999).
9. Choe, H. *et al.* Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120. *Cell* **114**, 161–170 (2003).
10. Fiete, D., Srivastava, V., Hindsgaul, O. & Baenziger, J.U. A hepatic reticuloendothelial cell receptor specific for SO4-4GalNAcβ 1,4GlcNAcβ 1,2Manα that mediates rapid clearance of lutropin. *Cell* **67**, 1103–1110 (1991).
11. Hanin, M. *et al.* Sulphation of *Rhizobium sp.* NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. *Mol. Microbiol.* **24**, 1119–1129 (1997).
12. Roche, P. *et al.* Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. *Cell* **67**, 1131–1143 (1991).
13. Shen, Y., Sharma, P., da Silva, F.G. & Ronald, P. The *Xanthomonas oryzae* pv. lozengeoryzae *raxP* and *raxQ* genes encode an ATP sulphurylase and adenosine-5′-phosphosulphate kinase that are required for AvrXa21 avirulence activity. *Mol. Microbiol.* **44**, 37–48 (2002).
14. Mougous, J.D. *et al.* Discovery of sulfated metabolites in mycobacteria with a genetic and mass spectrometric approach. *Proc. Natl. Acad. Sci. USA* **99**, 17037–17042 (2002).
15. McCarthy, C. Synthesis and release of sulfolipid by *Mycobacterium avium* during growth and cell division. *Infect. Immun.* **14**, 1241–1252 (1976).
16. Khoo, K.H. *et al.* Altered expression profile of the surface glycopeptidolipids in drug-resistant clinical isolates of *Mycobacterium avium* complex. *J. Biol. Chem.* **274**, 9778–9785 (1999).
17. Lopez Marin, L.M. *et al.* Structure of a novel sulfate-containing mycobacterial glycolipid. *Biochemistry* **31**, 11106–11111 (1992).
18. Goren, M.B., Brokl, O. & Schaefer, W.B. Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: correlation of virulence with elaboration of sulfatides and strongly acidic lipids. *Infect. Immun.* **9**, 142–149 (1974).
19. Goren, M.B. Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. I. Purification and properties. *Biochim. Biophys. Acta* **210**, 116–126 (1970).
20. Daffe, M. & Draper, P. The envelope layers of mycobacteria with reference to their pathogenicity. *Adv. Microb. Physiol.* **39**, 131–203 (1998).
21. Zhang, L., English, D. & Andersen, B.R. Activation of human neutrophils by *Mycobacterium tuberculosis*-derived sulfolipid-1. *J. Immunol.* **146**, 2730–2736 (1991).
22. Zhang, L., Gay, J.C., English, D. & Andersen, B.R. Neutrophil priming mechanisms of sulfolipid-I and *N*-formyl-methionyl-leucyl-phenylalanine. *J. Biomed. Sci.* **1**, 253–262 (1994).
23. Zhang, L., Goren, M.B., Holzer, T.J. & Andersen, B.R. Effect of *Mycobacterium tuberculosis*-derived sulfolipid I on human phagocytic cells. *Infect. Immun.* **56**, 2876–2883 (1988).
24. Julian, E. *et al.* Serodiagnosis of tuberculosis: comparison of immunoglobulin A (IgA) response to sulfolipid I with IgG and IgM responses to 2,3-diacyltrehalose, 2,3,6-triacyltrehalose, and cord factor antigens. *J. Clin. Microbiol.* **40**, 3782–3788 (2002).
25. Julian, E., Matas, L., Alcaide, J. & Luquin, M. Comparison of antibody responses to a potential combination of specific glycolipids and proteins for test sensitivity improvement in tuberculosis serodiagnosis. *Clin. Diagn. Lab. Immunol.* **11**, 70–76 (2004).
26. Converse, S.E. *et al.* MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc. Natl. Acad. Sci. USA* **100**, 6121–6126 (2003).
27. Gilleron, M. *et al.* Diacylated sulfoglycolipids are novel mycobacterial antigens stimulating CD1-restricted T cells during infection with *Mycobacterium tuberculosis*. *J. Exp. Med.* **199**, 649–659 (2004).
28. Sirakova, T.D., Thirumala, A.K., Dubey, V.S., Sprecher, H. & Kolattukudy, P.E. The *Mycobacterium tuberculosis pks2* gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276**, 16833–16839 (2001).
29. Domenech, P. *et al.* The role of MmpL8 in sulfatide biogensis and virulence of *Mycobacterium tuberculosis*. *J. Biol. Chem.* **279**, 21257–21265 (2004).
30. Rousseau, C. *et al.* Sulfolipid deficiency does not affect the virulence of *Mycobacterium tuberculosis* H37Rv in mice and guinea pigs. *Infect. Immun.* **71**, 4684–4690 (2003).
31. Cronan, G.E. & Keating, D.H. *Sinorhizobium meliloti* sulfotransferase that modifies lipopolysaccharide. *J. Bacteriol.* **186**, 4168–4176 (2004).
32. Cole, S.T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
33. Kakuta, Y., Pedersen, L.G., Pedersen, L.C. & Negishi, M. Conserved structural motifs in the sulfotransferase family. *Trends Biochem. Sci.* **23**, 129–130 (1998).
34. Elbein, A.D. & Mitchell, M. Levels of glycogen and trehalose in *Mycobacterium smegmatis* and the purification and properties of the glycogen synthetase. *J. Bacteriol.* **113**, 863–873 (1973).
35. Petzold, C.J., Leavell, M.D. & Leary, J.A. Screening and identification of acidic carbohydrates in bovine colostrum by using ion/molecule reactions and Fourier transform ion cyclotron resonance mass spectrometry: specificity toward phosphorylated complexes. *Anal. Chem.* **76**, 203–210 (2004).
36. Negishi, M. *et al.* Structure and function of sulfotransferases. *Arch. Biochem. Biophys.* **390**, 149–157 (2001).
37. Edavettal, S.C. *et al.* Crystal structure and mutational analysis of heparan sulfate 3-O-sulfotransferase isoform 1. *J. Biol. Chem.* **279**, 25789–25797 (2004).
38. Kakuta, Y., Sueyoshi, T., Negishi, M. & Pedersen, L.C. Crystal structure of the sulfotransferase domain of human heparan sulfate *N*-deacetylase/ *N*-sulfotransferase 1. *J. Biol. Chem.* **274**, 10673–10676 (1999).
39. Kakuta, Y., Pedersen, L.G., Carter, C.W., Negishi, M. & Pedersen, L.C. Crystal structure of estrogen sulphotransferase. *Nat. Struct. Biol.* **4**, 904–908 (1997).
40. Petrotchenko, E.V., Pedersen, L.C., Borchers, C.H., Tomer, K.B. & Negishi, M. The dimerization motif of cytosolic sulfotransferases. *FEBS Lett.* **490**, 39–43 (2001).
41. Goodsell, D.S. & Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
42. Chapman, E., Bryan, M.C. & Wong, C.H. Mechanistic studies of β-arylsulfotransferase IV. *Proc. Natl. Acad. Sci. USA* **100**, 910-5 (2003).
43. Kakuta, Y., Petrotchenko, E.V., Pedersen, L.C. & Negishi, M. The sulfuryl transfer mechanism. Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *J. Biol. Chem.* **273**, 27325–27330 (1998).
44. Chapman, E., Bryan, M.C. & Wong, C.H. Mechanistic studies of β-arylsulfotransferase IV. *Proc. Natl. Acad. Sci. USA* **100**, 910–915 (2003).
45. Pedersen, L.C., Petrotchenko, E., Shevtsov, S. & Negishi, M. Crystal structure of the human estrogen sulfotransferase-PAPS complex: evidence for catalytic role of Ser137 in the sulfuryl transfer reaction. *J. Biol. Chem.* **277**, 17928–17932 (2002).
46. Rini, J.M. Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577 (1995).
47. Elbein, A.D., Pan, Y.T., Pastuszak, I. & Carroll, D. New insights on trehalose: a multifunctional molecule. *Glycobiology* **13**, 17R–27R (2003).
48. Wolf, A., Kramer, R. & Morbach, S. Three pathways for trehalose metabolism in *Corynebacterium glutamicum* ATCC13032 and their significance in response to osmotic stress. *Mol. Microbiol.* **49**, 1119–1134 (2003).

© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol

49. Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. & Davies, G.J. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.* **9**, 1337–1346 (2002).
50. Ronning, D.R. *et al.* Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nat. Struct. Biol.* **7**, 141–146 (2000).
51. Armstrong, J.I. *et al.* A library approach to the generation of bisubstrate analogue sulfotransferase inhibitors. *Org. Lett.* **3**, 2657–2660 (2001).
52. Verdugo, D.E. *et al.* Discovery of estrogen sulfotransferase inhibitors from a purine library screen. *J. Med. Chem.* **44**, 2683–2686 (2001).
53. Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. & Jacobs, W.R. Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**, 1911–1919 (1990).
54. Parish, T. & Stoker, N.G. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA* plcABC mutant by gene replacement. *Microbiology* **146**, 1969–1975 (2000).
55. Kaps, I. *et al.* Energy transfer between fluorescent proteins using a co-expression system in *Mycobacterium smegmatis*. *Gene* **278**, 115–124 (2001).
56. Harth, G. & Horwitz, M.A. Expression and efficient export of enzymatically active *Mycobacterium tuberculosis* glutamine synthetase in *Mycobacterium smegmatis* and evidence that the information for export is contained within the protein. *J. Biol. Chem.* **272**, 22728–22735 (1997).
57. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
58. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. A* **276**, 307–326 (1997).
59. Terwilliger, T.C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
60. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
61. Holton, J. & Alber, T. Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci. USA* **101**, 1537–1542 (2004).
62. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
63. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
64. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
65. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129–147 (1993).
66. Ehrhardt, D.W. *et al. In vitro* sulfotransferase activity of NodH, a nodulation protein of *Rhizobium meliloti* required for host-specific nodulation. *J. Bacteriol.* **177**, 6237–6245 (1995).
67. Pratt, M.R., Leigh, C.D. & Bertozzi, C.R. Formation of 1,1-α,α-glycosidic bonds by intramolecular aglycone delivery. A convergent synthesis of trehalose. *Org. Lett.* **5**, 3185–3188 (2003).
68. Langston, S., Bernet, B. & Vasella, A. Temporary protection and activation in the regioselective synthesis of saccharide sulfates. *Helv. Chim. Acta* **77**, 2341–2353 (1994).
69. Goren, M.B. Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. II. Structural studies. *Biochim. Biophys. Acta* **210**, 127–138 (1970).
70. Minnikin, D.E., Kremer, L., Dover, L.G. & Besra, G.S. The methyl-branched fortifications of *Mycobacterium tuberculosis*. *Chem. Biol.* **9**, 545–553 (2002).


© 2004 Nature Publishing Group http://www.nature.com/natstructmolbiol