Sulfotransferases and Sulfatases in Mycobacteria

Joseph D. Mougous, Richard E. Green, Spencer J. Williams, Steven E. Brenner, and Carolyn R. Bertozzi

Howard Hughes Medical Institute
Department of Chemistry
Department of Molecular and Cell Biology
Department of Plant and Microbial Biology
University of California, Berkeley
Berkeley, California 94720

Analysis of the genomes of Mycobacterium tuberculosis, Mycobacterium leprae, Mycobacterium smegmatis, and Mycobacterium avium has revealed a large family of genes homologous to known sulfotransferases. Despite reports detailing a suite of sulfated glycolipids in many mycobacteria, a corresponding family of sulfotransferase genes remains uncharacterized. Here, a sequence-based analysis of newly discovered mycobacterial sulfotransferase genes, named stf1-stf10, is presented. Interestingly, two sulfotransferase genes are highly similar to mammalian sulfotransferases, increasing the list of mycobacterial eukaryotic-like protein families. The sulfotransferases join an equally complex family of mycobacterial sulfatases: a large family of sulfatase genes has been found in all of the mycobacterial genomes examined. As sulfated molecules are common mediators of cell-cell interactions, the sulfotransferases and sulfatases may be involved in regulating host-pathogen interactions.

Introduction

Sulfated carbohydrates are widespread in nature, predominantly represented on cell surfaces and in the extracellular space [1]. Many of these sulfated molecules have been implicated as important mediators of extracellular traffic and cell-cell communication in humans [2, 3]. In contrast to the well-studied roles for these molecules in eukaryotic systems, the importance of sulfated sugars and their associated enzymes in bacteria remains relatively unexplored. The two classes of enzymes responsible for the introduction and removal of sulfate esters are the sulfotransferases and the sulfatases, respectively (Figure 1). Sulfotransferases catalyze the transfer of a sulfuryl group from an activated donor, 3'-phosphoadenosine-5'-phosphosulfate (PAPS), to a small molecule, a carbohydrate, or a tyrosine residue within a protein. Sulfatases, on the other hand, catalyze the hydrolysis of sulfate esters (or N-sulfates) to an alcohol (or amine) and free sulfate.

The human sulfotransferase gene family now comprises over thirty members. Golgi-resident sulfotransferases, including the carbohydrate and tyrosylprotein sulfotransferase subfamilies, are of particular biomedical interest [1, 4, 5]. In mammals, it is well established that the functional paradigm for sulfation in the extracellular environment is to modulate binding interactions between proteins and glycoconjugates [6, 7]. In contrast to the large number of sulfotransferases that have been characterized in mammals, there have been few such enzymes studied in bacteria. The role of sulfotransferases in rhizobia extends the theme of sulfation that has emerged from studies involving eukaryotic systems into the domain of Prokarya. Rhizobia are nitrogen-fixing bacteria that enter into a symbiotic relationship with a variety of legumes. The sulfotransferases, NodH and NoeE, catalyze the sulfation of secreted glycolipid root nodulation factors [8–10]. Genetic studies have shown that the sulfate group is an important determinant of host specificity. Mutant strains lacking the sulfotransferases involved in the biosynthesis of nodulation factors exhibit a host range distinct from wild-type.

Members of the mycobacteria genera, including the human pathogens Mycobacterium tuberculosis and Mycobacterium avium, produce numerous sulfated glycolipids [11, 12]. The sulfatides, a family of sulfated glycolipids based on a common trehalose-2-sulfate core and restricted to the Mycobacterium tuberculosis complex, are the best characterized. The structure of sulfotolipid-1 (SL-1), the most abundant sulfatide, was elucidated by Goren et al. and is given in Figure 2A [13]. Interest in these compounds stemmed from early work by Middlebrook and Goren that correlated Mycobacterium tuberculosis strain virulence with their abundance [14]. A second sulfated mycobacterial compound has been structurally characterized from Mycobacterium avium. A 4-O-sulfated 6-deoxytalose residue was found in the glycopeptidolipid (GPL) of an ethambutol-resistant Mycobacterium avium strain cultured from a patient with AIDS [15]. GPL sulfation was found upregulated after the strain had acquired drug resistance. GPL sulfation has also been found in Mycobacterium fortuitum, in this case at the 2-hydroxyl group of a 3,4-dimethyl rhamnose residue. In this example, sulfation appeared to be a constitutive modification [16]. The structures of the sulfated compounds from Mycobacterium avium and Mycobacterium fortuitum are shown in Figures 2B and 2C. The sulfotransferases that sulfate these compounds have not been identified. To date, only one mycobacterial gene, Rv1373 from Mycobacterium tuberculosis, has been identified and shown to encode a sulfotransferase; the endogenous substrate for this enzyme is unknown [17].

Many mycobacteria have long been noted to possess arylsulfatase activity [18]. Indeed, arylsulfatase activity has been used, along with other standards, in taxonomic studies to identify and classify members of the genus [19]. Arylsulfatase activity has been detected in crude extracts of Mycobacterium marinum (formerly Mycobacterium piscium), Mycobacterium smegmatis, Mycobacterium avium, and a number of other mycobacteria, indicating that sulfatases are relatively widespread amongst this genus [20, 21]. Mycobacterial arylsulfatases have only been partially purified in a few cases; however, in no case has the corresponding sulfatase gene been identified. Several ORFs encoding putative sulfatases have
given that this organism also appears to lack ATP sulfurylase and APS kinase, two enzymes required for the biosynthesis of the sulfate donor, PAPS. The discovery of this putative mycobacterial sulfotransferase family is supported by the biochemical data (discussed above) that details the existence of numerous sulfated compounds in the genera.

Our initial pairwise sequence similarity search used the human carbohydrate sulfotransferase, GlcNAc-6-sulfotransferase GST3 [6]. The search revealed two M. tuberculosis ORFs, Rv2267c and Rv3529c, that encode putative sulfotransferases. A third ORF in M. tuberculosis, Rv1373, was already annotated as a putative sulfotransferase but shows little similarity to GST3 or the other mycobacterial sulfotransferases [25]. Rivera-Marrero et al. recently reported that Rv1373, heterologously expressed in E. coli, possesses sulfotransferase activity against human sulfatides (galactose, glucose, and lactose ceramide) [17]. We identified eight M. avium and one M. smegmatis ORF (see Supplemental Data) that had significant similarity to M. tuberculosis ORFs Rv2267c and Rv3529c. Table 1 summarizes these findings and introduces the nomenclature we assigned to the sequences.

To confirm the result suggested by pairwise alignments, we compared the Stf sequences to the Pfam Profile Hidden Markov Model of the sulfotransferase family [26]. For this we used the hmmssearch tool from the HMMER package [27, 28]. Profile Hidden Markov Models are more sensitive in detecting remote homologs than pairwise alignment methods [29]. The Pfam sulfotransferase Profile Hidden Markov Model is derived from a structurally informed alignment of 119 known and predicted sulfotransferases. As shown in Table 1, the E-values (expectation values) generated by these sequences ranged from highly (7.1E-67) to marginally (1.5E-2) significant. Interestingly, the Stf member that best fits the Pfam sulfotransferase Profile Hidden Markov Model, Stf2, bears the least similarity to the other putative mycobacterial sulfotransferases. Indeed, the significance of this match explains why only this gene was previously identified and annotated as a probable sulfotransferase. Stf2 is also unique in being most similar

Sulfotransferases in Mycobacteria

We analyzed the available sequenced genomes of mycobacteria and identified a family of genes with significant similarity to known sulfotransferases (see Supplemental Data). Putative sulfotransferases were initially identified in M. tuberculosis [22], and further analysis revealed similar open reading frames (ORFs) in M. smegmatis [23] and M. avium [23]. No putative sulfotransferase ORFs or pseudogenes were found in the decayed genome of M. leprae [24]. This is not surprising, however,
to eukaryotic cytosolic sulfotransferases. For this reason, Stf2 was chosen as the outgroup in the phylogenetic analysis (see below). In contrast to Stf2, the remaining Stf sequences are more similar to Golgi-resident sulfotransferases, with Stf3 being the most similar. This similarity is depicted in Figure 3A, using dot plots of Stf2 and 3 versus murine cytosolic estrogen sulfotransferase [30] and murine Golgi-resident carbohydrate sulfotransferase GST3 [6], respectively [31]. In the figure, arrows are used to highlight conserved sequence motifs unique to sulfotransferases.

After identifying the putative mycobacterial sulfotransferases, we further defined the relationships among sequences within the family using comprehensive pairwise analyses. Figure 3B shows the results of sequence alignments between each pair of Stf sequences. Analysis of the alignments revealed three regions of high conservation within and across species (Figure 3A). As discussed below, two of these regions are motifs involved in PAPS binding. To determine the evolutionary relationship between the genes, we built a phylogenetic tree using a composite alignment of the three conserved regions. Figure 4 shows that there is a single clade of apparent sulfotransferase orthologs with a representative from each species, which we named Stf1. The outgroup M. tuberculosis sequence was designated Stf2. The remaining sequences, one from M. tuberculosis and seven from M. avium, were designated Stf3 through Stf10. The number and heterogeneity of M. avium sulfotransferases suggests that the organism may be capable of producing a large repertoire of unique sulfated structures, consistent with previous findings [32].

Sulfotransferases invariably contain two highly conserved substrate binding regions termed the 5′-phosphosulfate binding loop (5′PSB) and the 3′-phosphate binding domain (3′PB). No other enzymes, including those known to bind PAPS, such as PAPS reductase and APS kinase, contain these motifs [33, 34]. Thus, their presence is highly indicative of a sulfotransferase gene. Both PAPS binding motifs can be identified in each member of the Stf gene family. Furthermore, the position of the motifs in the linear Stf sequences is consistent with other sulfotransferases (Figure 3B). Sequence logos [35] of the 5′PSB and 3′PB motifs of the Stf family versus the Pfam sulfotransferases are shown in Figure 5. As would be expected for a family of sulfotransferases, the 5′PSB and 3′PB motifs comprise the most highly conserved residues among the Stf sequences: 66% identity within the regions shown in Figure 5 as compared to 45% overall.

Existing structural and mutagenesis data for various sulfotransferases have implicated specific residues in the 5′PSB and 3′PB motifs as key contributors to substrate binding or catalysis. At the fourth position of the 5′PSB, for example, a basic residue is thought to be required as a proton donor to the product, 5′-phosphoadenosine-3′-phosphate, following transfer of the sulfonyl group [36–38]. Importantly, only basic residues are found at this position in the Stf family. Several crystal structures of sulfotransferases have shown important hydrogen bonding interactions between the 5′-phosphate of PAPS and residues 7 and 8 of the 5′PSB motif [37, 39–41]. In the crystal structure of heparin-N-sulfotransferase, the side chains of two threonine residues at these positions serve as hydrogen bond donors to the 5′-phosphate oxygen atoms. The presence of Thr at these positions seems to be a general phenomenon, as evidenced by the Pfam-derived sequence logo. Each Stf sequence, with the exception of Stf2, has Thr at both of these positions (Figure 3B; 5′PSB). It is noteworthy, however, that Stf2 does maintain the more highly conserved Thr at position 7.

Together, these data suggest that the identified sequences constitute a family of mycobacterial sulfotransferases. Determining which sulfotransferases are responsible for producing which specific sulfated compounds found in mycobacteria is the next step for researchers faced with the challenge of understanding the biological significance of these molecules. The volume of literature regarding the sulfolipid SL-1 makes the sulfotransferase(s) involved in its biosynthesis a particularly interesting target for study.

### Sulfotransferase Genes

As mentioned above, Stf2 and Stf3 return E-values of highest significance to eukaryotic sulfotransferases when evaluated against the GenBank nonredundant database (Table 1). Although biochemical characterization is still required, this finding increases the number of predominantly eukaryotic-like gene families in myco-

<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>Location¹/ORF No.</th>
<th>Pfam HMM Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>stf1</td>
<td>M. tuberculosis</td>
<td>Rv3529c</td>
<td>1.7E-03</td>
</tr>
<tr>
<td>stf2</td>
<td>M. avium</td>
<td>62: 28616-29593</td>
<td>5.5E-04</td>
</tr>
<tr>
<td>stf3</td>
<td>M. smegmatis</td>
<td>3269: 488152-487010</td>
<td>7.0E-03</td>
</tr>
<tr>
<td>stf4</td>
<td>M. tuberculosis</td>
<td>Rv1373</td>
<td>7.1E-07</td>
</tr>
<tr>
<td>stf5</td>
<td>M. avium</td>
<td>4: 40694-39531</td>
<td>4.0E-03</td>
</tr>
<tr>
<td>stf6</td>
<td>M. avium</td>
<td>16: 10074-11219</td>
<td>9.4E-06</td>
</tr>
<tr>
<td>stf7</td>
<td>M. avium</td>
<td>144: 9328-10719</td>
<td>3.2E-04</td>
</tr>
<tr>
<td>stf8</td>
<td>M. avium</td>
<td>93: 23510-24655</td>
<td>1.5E-02</td>
</tr>
<tr>
<td>stf9</td>
<td>M. avium</td>
<td>131: 7269-6100</td>
<td>5.6E-05</td>
</tr>
<tr>
<td>stf10</td>
<td>M. avium</td>
<td>130: 23247-22063</td>
<td>1.3E-05</td>
</tr>
<tr>
<td>stf11</td>
<td>M. avium</td>
<td>304: 13367-12225</td>
<td>8.8E-06</td>
</tr>
</tbody>
</table>

¹/ORF location given as contig number: nucleotide range in the unannotated genomes of M. smegmatis and M. avium.

Orthologs of M. tuberculosis stf1 are present in M. avium and M. smegmatis (Figure 4).
Figure 3. Blast E-Value and Dot-Plot Representation of Pairwise Stf Alignments and Stf Alignments with Other Eukaryotic Sulfotransferases

(A) Eukaryotic-like Stf sequences Stf2 and Stf3 are shown in dot-plot representation versus murine estrogen sulfotransferase (murine EST; GenBank accession number 12963514) and murine GST4 (GenBank accession number 6753419), respectively. For reference, a representative Stf versus Stf dot plot from (B) is also shown. E-values for each alignment are given. PAPS binding motifs are highlighted with arrows: 5'PSB, 5'-phosphosulfate binding motif (in red); 3'PB, 3'-phosphate binding motif (in blue).

(B) Matrix showing dot plots of Stf pairwise alignments with corresponding E-values.
bacteria to three. The two other eukaryotic-like gene families in \textit{M. tuberculosis} are the adenylyl cyclases \cite{42–44} and the Ser/Thr protein kinases \cite{45, 46}. In mammals, both of these gene families encode mediators of signaling cascades. In the extracellular matrix of mammalian cells, sulfate modification of proteins and carbohydrates is a crucial regulatory modification; therefore, the discovery of eukaryotic-like sulfotransferases in \textit{M. tuberculosis} may represent another example whereby this pathogen has adapted to its eukaryotic host by acquiring or evolving a similar signaling pathway. The observations that sulfotransferase genes are absent from most prokaryotic genomes and Stf family members resemble eukaryotic sulfotransferases suggest that the \textit{stf} genes may have been laterally transferred to mycobacteria.

**Sulfatases in Mycobacteria**

Several putative sulfatase sequences were identified in the original annotation of the genomes of \textit{M. tuberculosis} and \textit{M. leprae} (Figure 6) \cite{22, 24, 25}. Cole et al. found that in contrast to the sequence of the \textit{M. tuberculosis} genome, which contains 3924 genes, the \textit{M. leprae} genome encodes only approximately 1604 proteins and 1116 pseudogenes\cite{26}. In the case of \textit{M. leprae}, the putative sulfatases appear to have decayed into non-functional pseudogenes interrupted by frameshifts and stop codons \cite{24, 25}. Nonetheless, significant sequence similarity to genes in other mycobacteria is apparent, permitting the classification of the pseudogenes in \textit{M. leprae}. As can be seen in Figure 6A, sulfatases so far appear ubiquitous among the mycobacteria. \textit{M. tuberculosis} possesses six intact putative sulfatase genes, \textit{M. smegmatis} and \textit{M. avium} each contain three putative sulfatases, and \textit{M. leprae} possesses three sulfatase pseudogenes. Orthologs of the \textit{M. tuberculosis} and \textit{M. leprae} \textit{AtsA} and \textit{AtsG} were found in all the organisms examined, whereas orthologs of \textit{AtsF} were found only in \textit{M. tuberculosis} and \textit{M. smegmatis}, and an ortholog of the \textit{M. tuberculosis} \textit{AtsD} could only be found in \textit{M. leprae}. A substantially divergent sulfatase was detected in \textit{M. avium} (coined \textit{AtsI}) that aligned strongly with the well-studied sulfatase \textit{AtsA} from \textit{P. aeruginosa}.

To date, no study has described the cloning and biochemical characterization of any mycobacterial sulfatase. However, by sequence alignment we show that the mycobacterial \textit{AtsG} genes have greatest similarity to murine and human sulfamidas; \textit{M. tuberculosis} \textit{AtsG} shares 24\% identity with the murine sulfamidase (Figure 6). Sulfamidas catalyze the hydrolysis of the N-sulfate group, a functional group that has been identified in only two structures, heparin and heparan sulfate (sulfated glycosaminoglycans), both produced exclusively by multicellular eukaryotes. In mammals, this enzyme is a resident of the lysosome and forms part of a complex responsible for the degradation of heparin \cite{47}. A single bacterial sulfamidase has been isolated from \textit{Flavobacterium heparinum} \cite{48–50}, which forms part of a complex that can effect the complete degradation of heparin. However, cloning of the corresponding gene has not been reported. Nonetheless, the presence of putative sulfamidas in addition to other putative sulfatases in the genome of every sequenced mycobacterium raises the prospect of a novel host-pathogen interaction. As
Figure 6. Sequence Alignment Representation of Domains Conserved between Putative Mycobacterial and Various Well-Characterized Sulfatases

(A) Selected portions of the sequence alignment of mycobacterial putative sulfatases with selected, well-characterized sulfatases. The domains indicated are the regions or portions of the regions shown schematically in (B). Musmus_sulf, sulfamidase from *Mus musculus* (GenBank accession number AG41945); Pseaer_AtsA, arylsulfatase from *Pseudomonas aeruginosa* (GenBank accession number CAA88421); Snnmel_BetC, choline sulfatase from *Rhizobium meliloti* (GenBank accession number NP_385055); Myctub_AtS (Rv0711), Myctub_AtS (Rv3299c), Myctub_AtS (Rv0663), Myctub_AtS (Rv3076), Myctub_AtS (Rv2976), putative sulfatases from *M. tuberculosis*; Myclep_AtS (ML1853), Myclep_AtS (ML0795), Myclep_AtS (ML2525), putative sulfatase pseudogenes from *M. lepromatosis*; Mycavi_AtS, Mycavi_AtS, Mycavi_AtS, putative sulfatases from *M. avium*; Mycsme_AtS, Mycsme_AtS, Mycsme_AtS, Mycsme_AtS, putative sulfatases from *M. smegmatis*.

(B) Domain structure of a representative sulfatase, *Pseudomonas aeruginosa* AtsA. The domains indicated are regions of conserved sequence across the set of sulfatases shown in (A). Numbers shown below indicate the approximate boundaries of each proposed domain.

mentioned earlier, adhesion phenomena are critical in the pathogenesis of many bacteria, including *M. tuberculosis* and *M. avium*. During the early stages of infection, *M. tuberculosis* must adhere to epithelial cells. Surface-bound heparin binding hemagglutinins have been identified in *M. tuberculosis*, *M. bovis* BCG, and *M. leprae* [51]. Mycobacterial sulfamidases and sulfatases may act on extracellular glycosaminoglycans, resulting in the remodeling of sulfosulfatases in the extracellular matrix and thereby modulate bacterial adhesion. In the case of *M. leprae*, in which the sulfatase genes are not functional, such an interaction is less likely to occur. This is not surprising, given the difference in the pathology of this organism from that of other disease-causing mycobacteria.

Other genes from mycobacteria have high levels of similarity to well-studied O-sulfatases. For example, the *M. avium* Atsl is 65% identical to the sulfatase AtsA from *P. aeruginosa*. Additionally, the *M. tuberculosis* and *M. smegmatis* AtsF genes have considerable similarity to BetC (20% identity between BetC and *M. tuberculosis* AtsF), a choline-O-sulfatase from *S. melliloti*. In this organism, glycine betaine is accumulated in response to osmotic stress and appears to act as an...
osmoprotectant [52]. The biosynthesis of glycine betaine proceeds from choline-O-sulfate, by way of choline, through the action of BetC. The role of AtsF in M. tuberculosis and M. smegmatis is unknown but could also be related to osmoprotection. Other roles for choline-O-sulfate in mycobacteria are possible; for example, choline-O-sulfate is produced by some plants, apparently as an osmoprotectant [53], and by fungi for the storage of sulfate [54].

High levels of conservation of residues that have been demonstrated to be important for function can be seen in the sequence alignments of the putative mycobacterial sulfatases. For the sake of brevity, we have chosen to present abbreviated sequence alignments derived as follows. In the complete multiple sequence alignment of these predicted protein sequences, four domains that each contained high levels of conservation across all of the sequences were noted. These four domains, labeled I–IV, are illustrated using the P. aeruginosa AtsA as an example (Figure 6B). Domain I contains the greatest similarity among these sequences. Domains II and III are each rather short but have considerable sequence similarity, and Domain IV has moderate sequence similarity that stretches over the C-terminal end of the protein. Residues that have been identified and characterized as important for function through mutagenesis and structural studies are highly conserved and can be found in each of the four domains (Figure 6).

Sulfatases contain a remarkable co-/post-translational modification, namely the oxidation of an active site cysteine or serine residue to formylglycine. This unusual residue is believed to be intimately involved in the catalytic mechanism, functioning in its hydrated form as the enzymic nucleophile that attacks the sulfur atom of the sulfate ester [55]. Indeed, a rare autosomal recessive disorder, multiple sulfatase deficiency, is characterized by the lack of this co-/post-translational modification in all sulfatases, and thus a complete loss of sulfatase activity [56]. In eukaryotes, the sequence determinants that direct this modification have been investigated in some detail, and in every case this modification requires the presence of the sequence (C/S)XPXR [57]. The cysteine or serine residue undergoing oxidation is the first residue in this sequence, and in all of the mycobacterial genes described here but one, M. tuberculosis AtsH (which has a lysine in place of arginine), this residue and the surrounding motif are completely conserved (however, the complete motif is not present in the three M. leprae pseudogenes).

X-ray structures for three sulfatases have been determined: two human sulfatases, cerebroside 3-sulfate sulfatase (arylsulfatase A) [55], and N-acetylgalactosamine-4-sulfate sulfatase (arylsulfatase B) [58], and a single prokaryotic sulfatase, AtsA from P. aeruginosa [59]. These studies have identified a conserved set of amino acid residues that comprise the enzyme active site. Using the numbering of the P. aeruginosa AtsA, these are D13, D14, R55, K113, H115, H211, D317, N318, and K375 (excluding the enzymic nucleophile formylglycine discussed above and in reference 55). Further, in the case of the human arylsulfatase A, sequential mutation of each of the corresponding residues to alanine results in marked decreases in the catalytic efficiency of this enzyme, consistent with these residues playing a role in catalysis [60]. With regard to the mycobacterial putative sulfatase sequences, there are high levels of conservation among each of these residues and, in particular, among residues that have been directly implicated in the catalytic mechanism [58]. In almost every member of the putative mycobacterial sulfatase family, these residues are conserved, suggesting that they are indeed functional sulfatases.

Next Step: Role of Sulfated Metabolites in Host-Pathogen Interactions

We have reported the discovery and sequence-based analysis of a novel family of mycobacterial sulfotransferase genes. The discovery of this gene family reconciles previous biochemical data which demonstrated that mycobacteria produce a large repertoire of sulfated compounds. We have also noted many interesting features of the large and varied set of sulfatases in mycobacteria. We are not aware of any bacteria, from the more than 50 sequenced genomes in the public database, that contain as many putative sulfotransferases and sulfatases as any of the sequenced mycobacteria. The identification of these two large gene families will aid in the investigation of the sulfate metabolism pathways used by these organisms and increase our understanding of the roles of sulfated molecules in the biology and pathlogy of mycobacteria.

Many studies have shown that the purpose of eukaryotic and also some homologous prokaryotic enzymes similar to those described here is to modulate extracellular processes, such as cell adhesion and receptor/ligand interactions, by the addition and removal of sulfate esters. The question of whether or not the mycobacterial enzymes function in an analogous manner remains unanswered. Interestingly, like the sulfated glycolipids from rhizobia, the sulfated compounds identified in mycobacteria are found outside the cell, where they may be poised for interactions with either the host or another mycobacterial cell.

In the 40 years since the discovery of the sulfatides of M. tuberculosis, many researchers have sought to define their function. Studies using purified SL-1 have suggested that the sulfatides possess an ability to modulate the activation state of neutrophils and macrophages [61, 62]. Further progress toward elucidating the role of sulfatides in vivo has been hampered by a lack of knowledge regarding the genes involved in their synthesis, the difficulty in working with the pathogenic M. tuberculosis strains needed for its production, and the relative instability of the isolated molecule. Recently, however, by deleting a region of a polyketide synthase gene (pks2) of M. tuberculosis H37Rv, Sirakova et al. have generated a mutant strain that lacks SL-1 [63]. This SL-1-deficient strain presents the first opportunity to study the function of sulfatides in a known genetic background. The identification of a sulfotransferase family may also aid in the study of SL-1. Previous examples have shown that the sulfation state of a molecule can have a dramatic impact on its activity [8, 9, 64]. By analogy, deletion of the SL-1 sulfotransferase could have a large impact on the properties of this poorly understood glycolipid. In this regard, enzymes that act upstream of the sulfotransferases may also prove valu-
able targets. For example, the enzymes involved in the biosynthesis of PAPS could be worthy targets of study; disruption of PAPS biosynthesis would be expected to completely arrest the formation of sulfated metabolites.

In eukaryotes, sulfatases are typically localized in the lysosome and are responsible for the recycling of sulfated structures such as sulfomucins and sulfated glycosaminoglycans [65]. By contrast, despite the long history of observations of sulfatase activity in crude extracts, the biological roles of most bacterial sulfatases remain uncertain. In a few cases, however, they have been implicated in important host-pathogen interactions. For example, in the case of colonic sulfomucins, removal of sulfate groups primes the mucins for degradation by secreted bacterial glycosidasases [66]. Thus, sulfate esters appear to protect the host, with elevated levels of bacterial mucin-degrading sulfatases being correlated with ulcerative colitis [67]. In another case, deletion of the sulfatase aslA from a cerebrospinal isolate of E. coli K1 (responsible for neonatal E. coli meningitis) reduces the ability of this organism to invade the brain microvasculature [68].

The role of sulfatases in the biology and pathology of mycobacteria has not been investigated. Pathogenic mycobacteria, including M. tuberculosis, M. leprae, and M. avium, produce on their surface a hemagglutinin (hbbH) that confers a high affinity for heparin and that is involved in adherence to epithelial cells [69–71]. During the early stages of infection, the M. tuberculosis bacillus must adhere to the target tissue prior to phagocytosis by the alveolar macrophages. It has been suggested that attachment of M. tuberculosis to epithelial cells and the extracellular matrix occurs through interactions with sulfated glycoconjugates [51]. In support of this, Pethé et al. showed that deletion of hbbHa in a pathogenic M. tuberculosis strain or the vaccine strain M. bovis BCG substantially reduced the ability of these enzymes to adhere to, invade, and survive within epithelial cells [51]. Whether sulfatases are involved in the remodeling of mucosal surfaces and modulation of the adherence of pathogenic mycobacteria has not been addressed. Mycobacterial sulfatases could also be involved in the acquisition of sulfur in the form of sulfate from the environment or from the human host. The intracellular environments are relatively nutrient poor, and these enzymes could provide alternate sources of sulfur for the production of essential amino acids and sulfur-containing cofactors. In this regard, enzymes that act upstream of the sulfatases may also prove valuable targets. Disruption of the aslA described enzyme(s) necessary for the posttranslational modification of cysteine to formylglycine represents an excellent approach to generating a complete knockout of sulfatase activity [72].

Chemistry & Biology invites your comments on this topic. Please write to the editors at chembio@cell.com.

Supplemental Data
Supplemental data, including methods and nucleotide and protein sequences of previously unannotated mycobacterial sulfotransferases and sulfatases discussed in this article, is available through the Chemistry & Biology production department. Please write to chembio@cell.com for a PDF.

Acknowledgments
The authors wish to thank Dr. Stefan Hemmerich of Thios Pharmaceuticals for his assistance in the identification of the sulfotransferases. J.D.M. is supported by a Ford Foundation Predoctoral Fellowship. S.J.W. is supported by a Howard Hughes Medical Institute Fellowship of the Life Sciences Research Foundation. Sequencing of M. avium and M. smegmatis was accomplished with support from the National Institute of Allergy and Infectious Diseases. Work in the authors’ laboratories was supported by grants to C.R.B. (AI51622) and S.E.B. (K22-HG00056) from the National Institutes of Health.

References
16. Lopez Marin, L.M., Laneelle, M.A., Prome, D., Laneelle, G.,


