Chemistry & Biology, Vol. 9, 767–776, July, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00175-8

Sulfotransferases and Sulfatases in Mycobacteria

Crosstalk

Joseph D. Mougous,[3] Richard E. Green,[3,4,6] Spencer J. Williams,[2,6] Steven E. Brenner,[3,4] and Carolyn R. Bertozzi[1,2,3,5]
[1]Howard Hughes Medical Institute
[2]Department of Chemistry
[3]Department of Molecular and Cell Biology
[4]Department of Plant and Microbial Biology
University of California, Berkeley
Berkeley, California 94720

Analysis of the genomes of *M. tuberculosis*, *M. leprae*, *M. smegmatis*, and *M. 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 *stf*1-*stf*10, 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 *M. tuberculosis* and *M. 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 *M. tuberculosis* complex, are the best characterized. The structure of sulfolipid-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 *M. tuberculosis* strain virulence with their abundance [14]. A second sulfated mycobacterial compound has been structurally characterized from *M. avium*. A 4-*O*-sulfated 6-deoxytalose residue was found in the glycopeptidolipid (GPL) of an ethambutol-resistant *M. 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 *M. avium* and *M. fortuitum* are shown in Figures 2B and 2C. The sulfotransferases that sulfate these compounds have not been identifed. To date, only one mycobacterial gene, Rv1373 from *M. 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*, *M. 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

[5]Correspondence: bertozzi@cchem.berkeley.edu
[6]These authors contributed equally to this work.

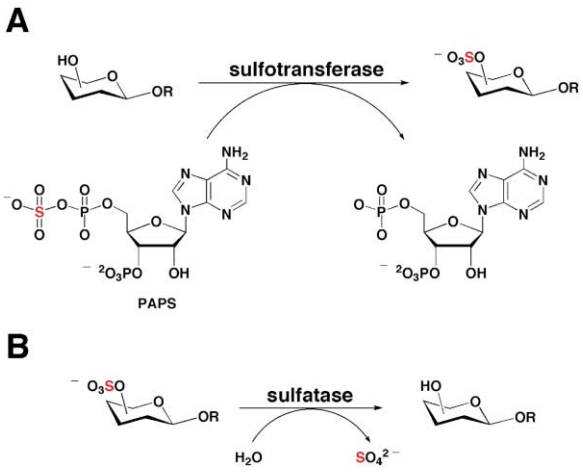


Figure 1. General Reactions Catalyzed by Sulfotransferases (A) and Sulfatases (B)

been annotated in the genome of *M. tuberculosis*, but their substrates and functions remain undefined.

To begin to address this issue, we have searched the sequenced genomes of mycobacteria for sulfotransferase and sulfatase genes. Our search led to the identification of an extensive family of sulfotransferases that are similar to known eukaryotic carbohydrate sulfotransferases. Additionally, a large family of sulfatase genes was found in each of the species analyzed. While some of these genes were similar to the annotated sulfatases in the genome of *M. tuberculosis*, several sulfatases novel to the mycobacterial genus were noted.

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, 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 hmmsearch 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

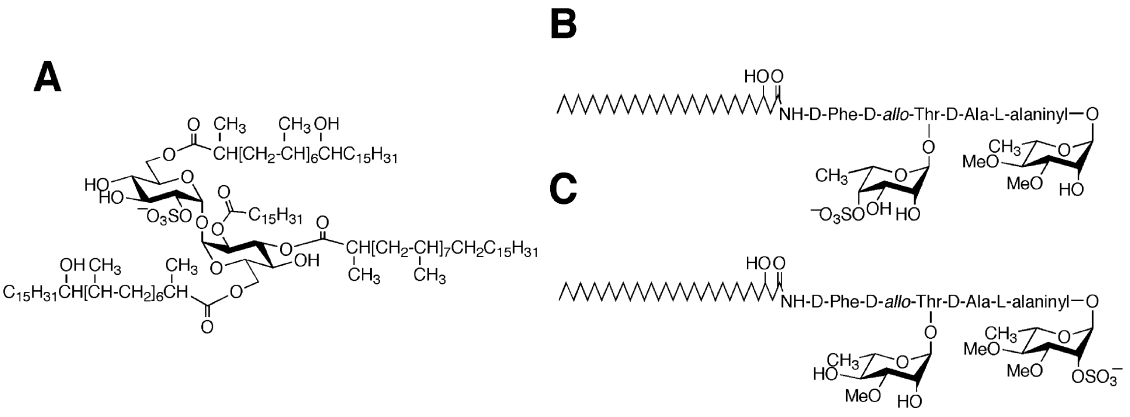


Figure 2. Structures of Sulfated Compounds in Mycobacteria that Have Been Structurally Elucidated

(A), SL-1 from *Mycobacterium tuberculosis*; (B), sulfated glycopeptidolipid from *Mycobacterium avium*; (C), sulfated glycopeptidolipid from *Mycobacterium fortuitum*.

Table 1. Annotation and Pfam Sulfotransferase HMM Score of the *stf* Gene Family

Name	Species	Location[1]/ORF No.	Pfam HMM Score
*stf*1[2]	*M. tuberculosis*	Rv3529c	1.7E-03
	M. avium	62: 28616-29593	5.5E-04
	M. smegmatis	3269: 488152-487010	7.0E-03
*stf*2	*M. tuberculosis*	Rv1373	7.1E-67
*stf*3	*M. tuberculosis*	Rv2267c	5.0E-07
*stf*4	*M. avium*	4: 40694-39531	4.0E-03
*stf*5	*M. avium*	16: 10074-11219	9.4E-06
*stf*6	*M. avium*	144: 9328-10719	3.2E-04
*stf*7	*M. avium*	93: 23510-24655	1.5E-02
*stf*8	*M. avium*	131: 7269-6100	5.6E-05
*stf*9	*M. avium*	130: 23247-22063	1.3E-05
*stf*10	*M. avium*	304: 13367-12225	8.8E-06

[1]ORF location given as contig number: nucleotide range in the unannotated genomes of *M. smegmatis* and *M. avium*.
[2]Orthologs of *M. tuberculosis stf*1 are present in *M. avium* and *M. smegmatis* (Figure 4).

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].

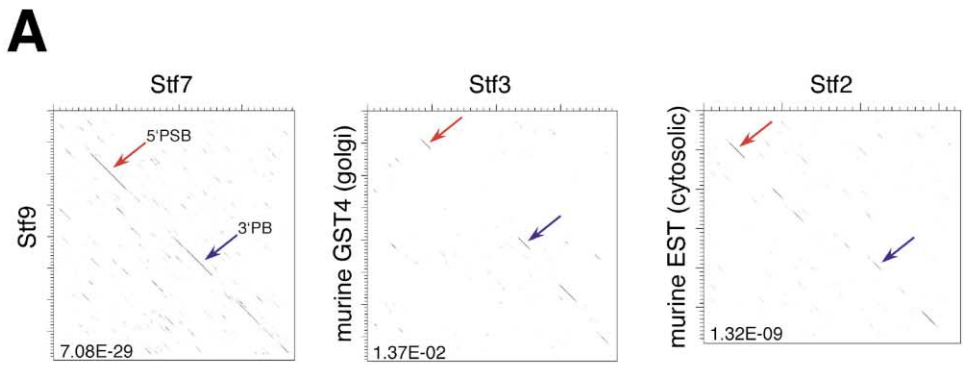
Sulfotransferases invariably contain two highly conserved substrate binding regions termed the 5′-phosphosulfate binding loop (5′PSB) and the 3′-phosphate binding domain (3′PB). No other enzymes, including those known to bind PAPS, such as PAPS reductase and APS kinase, contain these motifs [33, 34]. 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 sulfuryl 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.

Stf Family Members Resemble Eukaryotic 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-



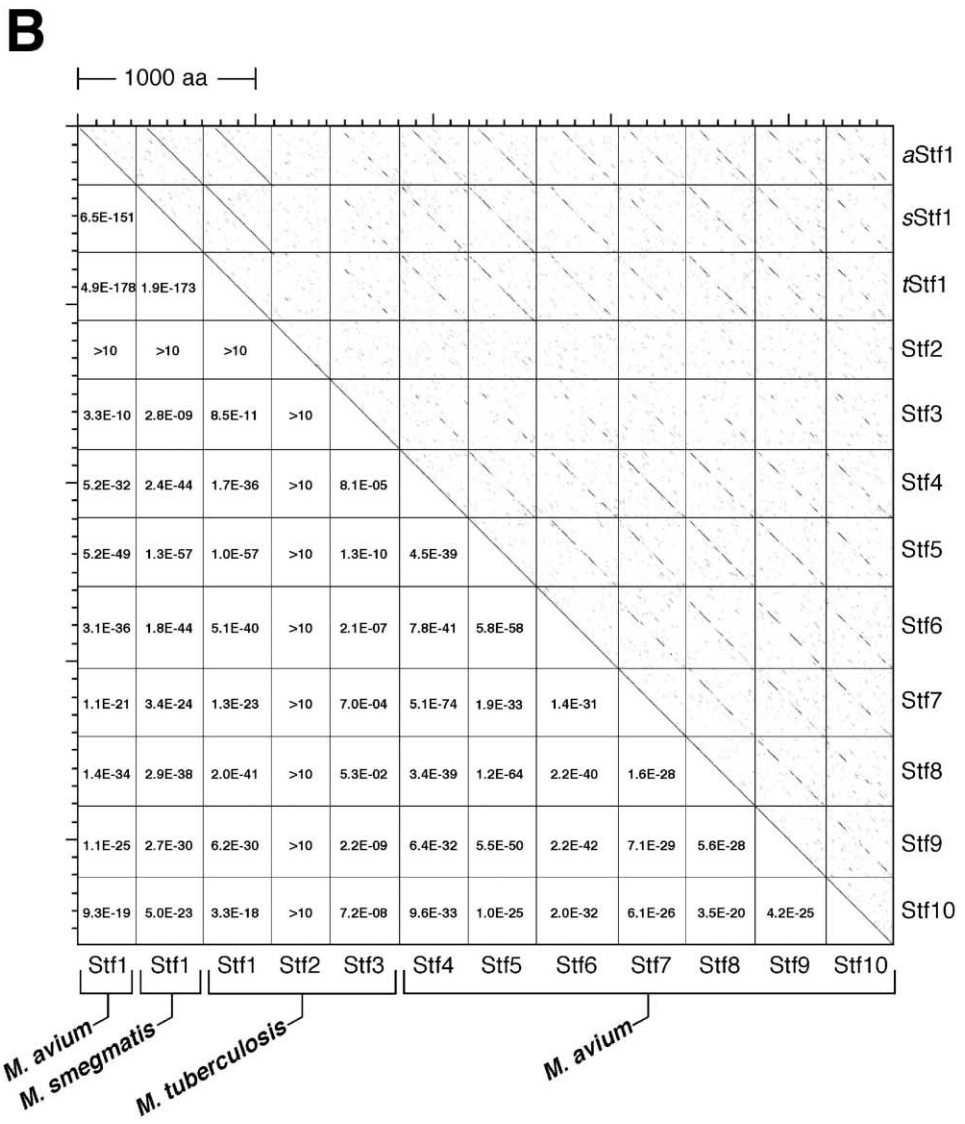


Figure 3. Blast E-Value and Dot-Plot Representation of Pairwise Stf Alignments and Stf Alignments with Other Eukaryotic Sulfotransferases

(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.

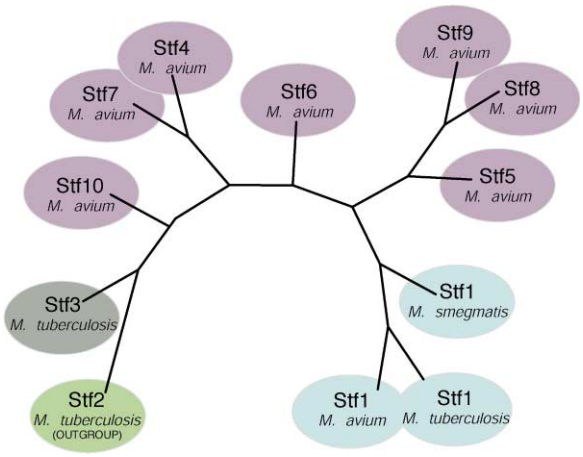


Figure 4. Unrooted Phylogenetic Tree of the Stf Family

Stf1 orthologs are shown in blue. The remaining *M. avium* family members are given in purple to highlight their clustered topology. The remaining *M. tuberculosis* Stf proteins are shown in gray and green (outgroup). Branch lengths between Stf family members are not meaningful (as described in the Methods section of the Supplemental Data).

bacteria to three. The two other eukaryotic-like gene families in *M. tuberculosis* are the adenylyl cyclases [42–44] and the Ser/Thr protein kinases [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 *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 *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 *M. tuberculosis* and *M. leprae* (Figure 6) [22, 24, 25]. Cole et al. found that in contrast to the sequence of the *M. tuberculosis* genome, which contains 3924 genes, the *M. leprae* genome encodes only approximately 1604 proteins and 1116 pseudogenes [25]. In the case of *M. leprae*, the putative sulfatases appear to have decayed into nonfunctional pseudogenes interrupted by frameshifts and stop codons [24, 25]. Nonetheless, significant sequence similarity to genes in other mycobacteria is apparent, permitting the classification of the pseudogenes in *M. leprae*. As can be seen in Figure 6A, sulfatases so far appear ubiquitous among the mycobacteria. *M. tuberculosis* possesses six intact putative sulfatase genes, *M. smegmatis* and *M. avium* each contain three putative sulfatases, and *M. leprae* possesses three sulfatase pseudogenes. Orthologs of the *M. tuberculosis* and *M. leprae* AtsA and AtsG were found in all the organisms examined, whereas orthologs of AtsF were found only in *M. tuberculosis* and *M. smegmatis*, and an ortholog of the *M. tuberculosis* AtsD could only be found in *M. leprae*. A substantially divergent sulfatase was detected in *M. avium* (coined AtsI) that aligned strongly with the well-studied sulfatase AtsA from *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 AtsG genes have greatest similarity to murine and human sulfamidases; *M. tuberculosis* AtsG shares 24% identity with the murine sulfamidase (Figure 6). Sulfamidases 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 [47]. A single bacterial sulfamidase has been isolated from *Flavobacterium heparinum* [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 sulfamidases in addition to other putative sulfatases in the genome of every sequenced mycobacterium raises the prospect of a novel host-pathogen interaction. As

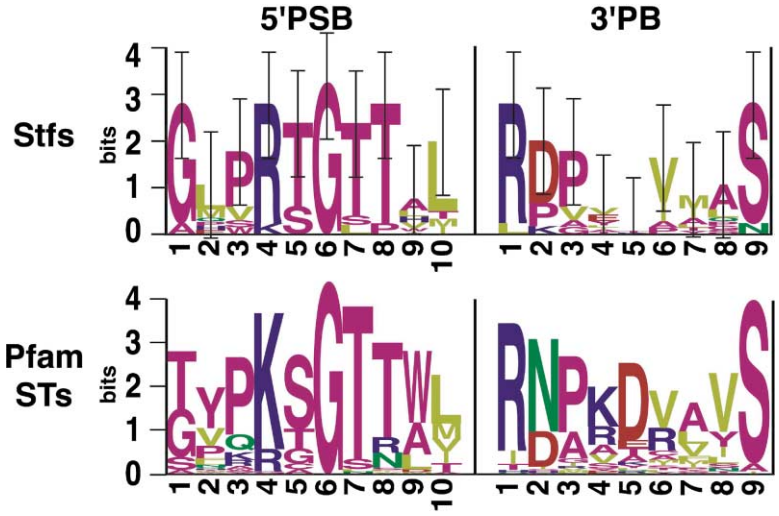


Figure 5. Sequence Logo Comparison of the 5′PSB and 3′PB Regions of the Stf and Pfam Sulfotransferase Sequences

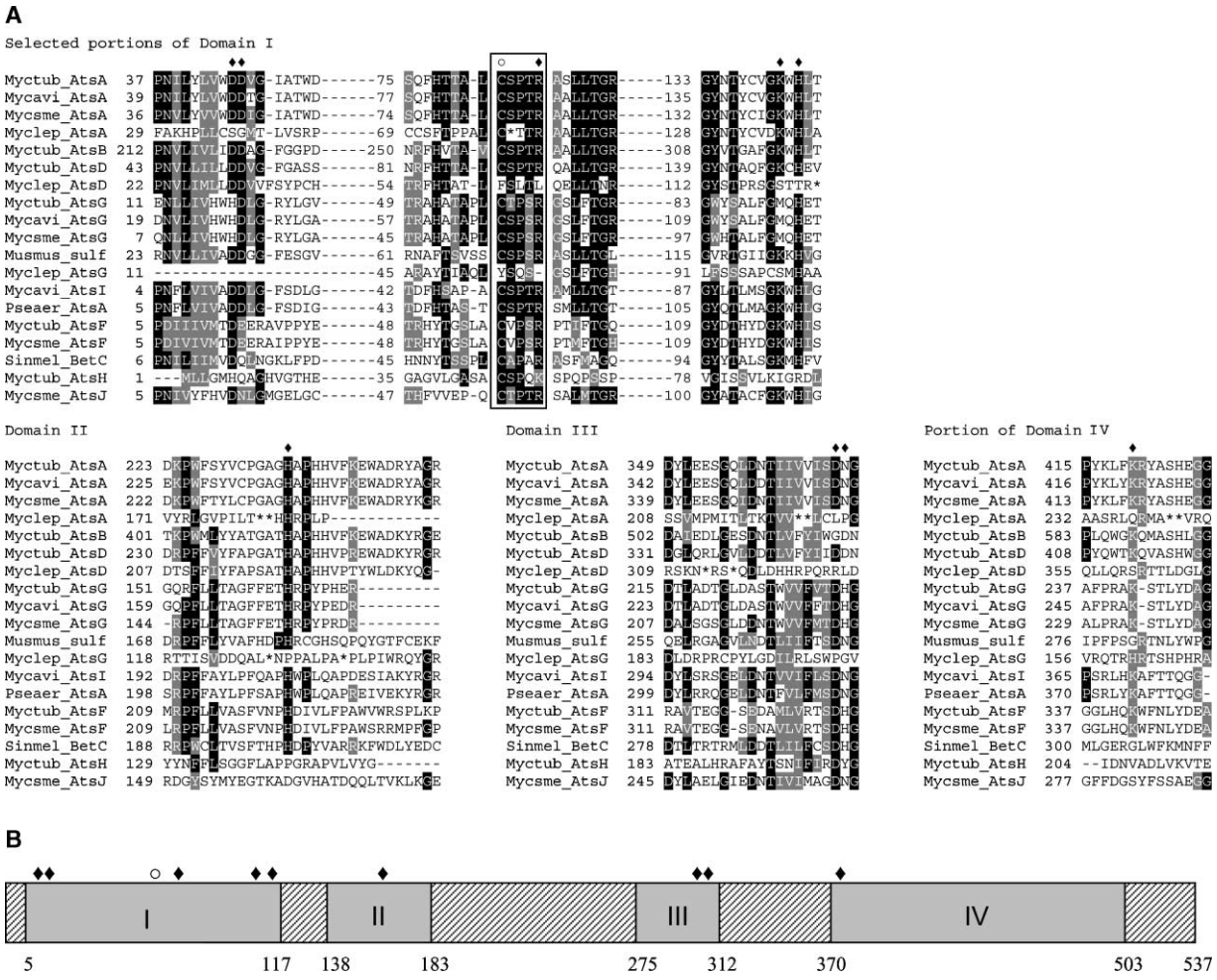


Figure 6. Sequence Alignment Representation of Domains Conserved between Putative Mycobacterial and Various Well-Characterized Sulfatases

(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); Sinmel_BetC, choline sulfatase from *Rhizobium meliloti* (GenBank accession number NP_385055); Myctub_AtsA (Rv0711), Myctub_AtsB (Rv3299c), Myctub_AtsD (Rv0663), Myctub_AtsF (Rv3076), Myctub_AtsG (Rv0296c), Myctub_AtsH (Rv3796), putative sulfatases from *M. tuberculosis*; Myclep_AtsA (ML1853), Myclep_AtsD (ML0795), Myclep_AtsG (ML2525), putative sulfatase pseudogenes from *M. leprae*; Mycavi_AtsA, Mycavi_AtsG, Mycavi_AtsI, putative sulfatases from *M. avium*; Mycsme_AtsA, Mycsme_AtsG, Mycsme_AtsF, Mycsme_AtsJ, putative sulfatases from *M. smegmatis*. Residues indicated with a diamond and open circle indicate conserved residues that comprise the active site and are important for activity as shown by site-directed mutagenesis. The boxed residues indicate the CXPXR motif required for co-/post-translational modification of sulfatases. The cysteine residue, indicated with an open circle, is the residue shown to be oxidized to formylglycine, an important modification for the proposed mechanism.

(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 sulfoforms 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* AtsI 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. meliloti*. 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 [59]. 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 pathology 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 glycosidases [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 (*hbh*A) 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, Pethe et al. showed that deletion of *hbhA* in a pathogenic *M. tuberculosis* strain or the vaccine strain *M. bovis* BCG substantially reduced the ability of these mutants 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 intra- and extracellular 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 as yet undescribed 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 chembiol@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 chembiol@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

1. Bowman, K.G., and Bertozzi, C.R. (1999). Carbohydrate sulfotransferases: mediators of extracellular communication. Chem. Biol. *6*, R9–R22.
2. Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. Glycobiology *3*, 97–130.
3. Hooper, L.V., Manzella, S.M., and Baenziger, J.U. (1996). From legumes to leukocytes: biological roles for sulfated carbohydrates. FASEB J. *10*, 1137–1146.
4. Kehoe, J.W., and Bertozzi, C.R. (2000). Tyrosine sulfation: a modulator of extracellular protein-protein interactions. Chem. Biol. *7*, R57–R61.
5. Armstrong, J.I., and Bertozzi, C.R. (2000). Sulfotransferases as targets for therapeutic intervention. Curr. Opin. Drug. Discov. Devel. *3*, 502–515.
6. Bistrup, A., Bhakta, S., Lee, J.K., Belov, Y.Y., Gunn, M.D., Zuo, F.R., Huang, C.C., Kannagi, R., Rosen, S.D., and Hemmerich, S. (1999). Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin. J. Cell Biol. *145*, 899–910.
7. Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N.P., Gerard, C., Sodroski, J., and Choe, H. (1999). Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. Cell *96*, 667–676.
8. Quesada-Vincens, D., Hanin, M., Broughton, W.J., and Jabbouri, S. (1998). In vitro sulfotransferase activity of NoeE, a nodulation protein of *Rhizobium* sp. NGR234. Mol. Plant Microbe Interact. *11*, 592–600.
9. Hanin, M., Jabbouri, S., Quesada-Vincens, D., Freiberg, C., Perret, X., Prome, J.C., Broughton, W.J., and Fellay, R. (1997). Sulphation of *Rhizobium sp*. NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. Mol. Microbiol. *24*, 1119–1129.
10. Cullimore, J.V., Ranjeva, R., and Bono, J.J. (2001). Perception of lipo-chitooligosaccharidic Nod factors in legumes. Trends Plant Sci. *6*, 24–30.
11. Tsukamura, M., and Mizuno, S. (1980). Subgrouping of strains of *Mycobacterium avium-Mycobacterium intracellulare* complex by thin-layer chromatography of acetone-soluble fraction after incubation with ^{35}S-sulfate. Kekkaku *55*, 481–484.
12. Tsukamura, M., Mizuno, S., and Toyama, H. (1984). Differentiation of mycobacterial species by investigation of petroleum ether-soluble sulfolipids using thin-layer chromatography after incubation with ^{35}S-sulfate. Microbiol. Immunol. *28*, 965–974.
13. Goren, M.B. (1970). Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. II. Structural studies. Biochim. Biophys. Acta *210*, 127–138.
14. Goren, M.B., Brokl, O., and Schaefer, W.B. (1974). 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.
15. Khoo, K.H., Jarboe, E., Barker, A., Torrelles, J., Kuo, C.W., and Chatterjee, D. (1999). Altered expression profile of the surface glycopeptidolipids in drug- resistant clinical isolates of *Mycobacterium avium* complex. J. Biol. Chem. *274*, 9778–9785.
16. Lopez Marin, L.M., Laneelle, M.A., Prome, D., Laneelle, G.,

Prome, J.C., and Daffe, M. (1992). Structure of a novel sulfate-containing mycobacterial glycolipid. Biochemistry *31*, 11106–11111.

17. Rivera-Marrero, C.A., Ritzenthaler, J.D., Newburn, S.A., Roman, J., and Cummings, R.D. (2002). Molecular cloning and expression of a novel glycolipid sulfotransferase in *Mycobacterium tuberculosis*. Microbiology *148*, 783–792.
18. Dodgson, K.S., White, G.F., and Fitzgerald, J.W. (1982). Sulfatases of Microbial Origin, Volume 2 (Boca Raton, FL: CRC Press).
19. Levy-Frebault, V.V., and Portaels, F. (1992). Proposed minimal standards for the genus *Mycobacterium* and for description of new slowly growing *Mycobacterium* species. Int. J. Syst. Bacteriol. *42*, 315–323.
20. Whitehead, J.E.M., Morrison, A.R., and Young, L. (1952). Bacterial arylsulphatase. Biochem. J. *51*, 585–594.
21. Falkinham, J.O., 3rd. (1990). Arylsulfatase activity of *Mycobacterium avium*, *M. intracellulare*, and *M. scrofulaceum*. Int. J. Syst. Bacteriol. *40*, 66–70.
22. http://genolist.pasteur.fr/TubercuList
23. http://www.tigr.org
24. http://genolist.pasteur.fr/Leproma
25. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., III, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature *393*, 537–544.
26. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L.E. (2002). The Pfam protein families database. Nucleic Acids Res. *30*, 276–280.
27. Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763.
28. http://hmmer.wustl.edu
29. Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. (1998). Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. J. Mol. Biol. *284*, 1201–1210.
30. Kakuta, Y., Pedersen, L.G., Pedersen, L.C., and Negishi, M. (1998). Conserved structural motifs in the sulfotransferase family. Trends Biochem. Sci. *23*, 129–130.
31. Hemmerich, S., and Rosen, S.D. (2000). Carbohydrate sulfotransferases in lymphocyte homing. Glycobiology *10*, 849–856.
32. McCarthy, C. (1976). Synthesis and release of sulfolipid by *Mycobacterium avium* during growth andcell division. Infect. Immun. *14*, 1241–1252.
33. MacRae, I.J., Segel, I.H., and Fisher, A.J. (2000). Crystal structure of adenosine 5′-phosphosulfate kinase from *Penicillium chrysogenum*. Biochemistry *39*, 1613–1621.
34. Savage, H., Montoya, G., Svensson, C., Schwenn, J.D., and Sinning, I. (1997). Crystal structure of phosphoadenylyl sulphate (PAPS) reductase: a new family of adenine nucleotide alpha hydrolases. Structure *5*, 895–906.
35. Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. *18*, 6097–6100.
36. Marsolais, F., Laviolette, M., Kakuta, Y., Negishi, M., Pedersen, L.C., Auger, M., and Varin, L. (1999). 3′-Phosphoadenosine 5′-phosphosulfate binding site of flavonol 3- sulfotransferase studied by affinity chromatography and ^{31}P NMR. Biochemistry *38*, 4066–4071.
37. Kakuta, Y., Sueyoshi, T., Negishi, M., and Pedersen, L.C. (1999). Crystal structure of the sulfotransferase domain of human heparan sulfate N-deacetylase/N-sulfotransferase 1. J. Biol. Chem. *274*, 10673–10676.
38. Sueyoshi, T., Kakuta, Y., Pedersen, L.C., Wall, F.E., Pedersen, L.G., and Negishi, M. (1998). A role of Lys614 in the sulfotransferase activity of human heparan sulfate *N*-deacetylase/*N*-sulfotransferase. FEBS Lett. *433*, 211–214.
39. Kakuta, Y., Pedersen, L.G., Carter, C.W., Negishi, M., and Pedersen, L.C. (1997). Crystal structure of estrogen sulphotransferase. Nat. Struct. Biol. *4*, 904–908.
40. Komatsu, K., Driscoll, W.J., Koh, Y.C., and Strott, C.A. (1994). A P-loop related motif (GxxGxxK) highly conserved in sulfotransferases is required for binding the activated sulfate donor. Biochem. Biophys. Res. Commun. *204*, 1178–1185.
41. Pedersen, L.C., Petrotchenko, E.V., and Negishi, M. (2000). Crystal structure of SULT2A3, human hydroxysteroid sulfotransferase. FEBS Lett. *475*, 61–64.
42. Guo, Y.L., Seebacher, T., Kurz, U., Linder, J.U., and Schultz, J.E. (2001). Adenylyl cyclase Rv1625c of *Mycobacterium tuberculosis*: a progenitor of mammalian adenylyl cyclases. EMBO J. *20*, 3667–3675.
43. McCue, L.A., McDonough, K.A., and Lawrence, C.E. (2000). Functional classification of cNMP-binding proteins and nucleotide cyclases with implications for novel regulatory pathways in *Mycobacterium tuberculosis*. Genome Res. *10*, 204–219.
44. Reddy, S.K., Kamireddi, M., Dhanireddy, K., Young, L., Davis, A., and Reddy, P.T. (2001). Eukaryotic-like adenylyl cyclases in *Mycobacterium tuberculosis* H37Rv. Cloning and characterization. J. Biol. Chem. *276*, 35141–35149.
45. Av-Gay, Y., Jamil, S., and Drews, S.J. (1999). Expression and characterization of the *Mycobacterium tuberculosis* serine/threonine protein kinase PknB. Infect. Immun. *67*, 5676–5682.
46. Av-Gay, Y., and Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. Trends Microbiol. *8*, 238–244.
47. Anson, D.S., and Bielicki, J. (1999). Sulphamidase. Int. J. Biochem. Cell Biol. *31*, 363–367.
48. Dietrich, C.P. (1969). Enzymic degradation of heparin. A sulphamidase and a sulphoesterase from *Flavobacterium heparinum*. Biochem. J. *111*, 91–95.
49. McLean, M.W., Bruce, J.S., Long, W.F., and Williamson, F.B. (1984). *Flavobacterium heparinum* 2-O-sulphatase for 2-O-sulphato-delta 4,5- glycuronate-terminated oligosaccharides from heparin. Eur. J. Biochem. *145*, 607–615.
50. Bruce, J.S., McLean, M.W., Long, W.F., and Williamson, F.B. (1987). *Flavobacterium heparinum* sulphamidase for D-glucosamine sulphamate. Purification and characterisation. Eur. J. Biochem. *165*, 633–638.
51. Pethe, K., Alonso, S., Biet, F., Delogu, G., Brennan, M.J., Locht, C., and Menozzi, F.D. (2001). The heparin-binding haemagglutinin of *M. tuberculosis* is required for extrapulmonary dissemination. Nature *412*, 190–194.
52. Østerås, M., Boncompagni, E., Vincent, N., Poggi, M.-C., and Le Rudulier, D. (1998). Presence of a gene encoding choline sulfatase in *Sinorhizobium meliloti* bet operon: choline-*O*-sulfate is metabolized into glycine betaine. Proc. Natl. Acad. Sci. USA *95*, 11394–11399.
53. Nissen, P., and Benson, A.A. (1961). Choline sulfate in higher plants. Science *134*, 1759.
54. Spencer, B., Hussey, E.C., Orsi, B.A., and Scott, J.M. (1968). Mechanism of choline *O*-sulphate utilization in fungi. Biochem. J. *106*, 461–469.
55. Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. (1998). Crystal structure of human arylsulfatase A: The aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. Biochemistry *37*, 3654–3664.
56. Schmidt, B., Selmer, T., Ingendoh, A., and von Figura, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. Cell *82*, 271–278.
57. Dierks, T., Lecca, M.R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999). Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. EMBO J. *18*, 2084–2091.
58. Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997). Structure of a human lysosomal sulfatase. Structure *5*, 277–289.
59. Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M.A., and Uson, I. (2001). 1.3 A structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. Structure *9*, 483–491.
60. Waldow, A., Schmidt, B., Dierks, T., von Bulow, R., and von Figura, K. (1999). Amino acid residues forming the active site

of arylsulfatase A – Role in catalytic activity and substrate binding. J. Biol. Chem. *274*, 12284–12288.

61. Pabst, M.J., Gross, J.M., Brozna, J.P., and Goren, M.B. (1988). Inhibition of macrophage priming by sulfatide from *Mycobacterium tuberculosis*. J. Immunol. *140*, 634–640.
62. Brozna, J.P., Horan, M., Rademacher, J.M., Pabst, K.M., and Pabst, M.J. (1991). Monocyte responses to sulfatide from *Mycobacterium tuberculosis*: inhibition of priming for enhanced release of superoxide, associated with increased secretion of interleukin-1 and tumor necrosis factor alpha, and altered protein phosphorylation. Infect. Immun. *59*, 2542–2548.
63. Sirakova, T.D., Thirumala, A.K., Dubey, V.S., Sprecher, H., and Kolattukudy, P.E. (2001). The *Mycobacterium tuberculosis pks2* gene encodes the synthase for the hepta- and octamethyl branched fatty acids required for sulfolipid synthesis. J. Biol. Chem. *23*, 16833–16839.
64. Roche, P., Debelle, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Denarie, J., and Prome, J.C. (1991). Molecular basis of symbiotic host specificity in Rhizobium meliloti: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. Cell *67*, 1131–1143.
65. Nieuw Amerongen, A.V., Bolscher, J.G.M., Bloemena, E., and Veerman, E.C.I. (1998). Sulfomucins in the human body. Biol. Chem. *379*, 1–18.
66. Tsai, H.H., Sunderland, D., Gibson, G.R., Hart, C.A., and Rhodes, J.M. (1992). A novel mucin sulphatase from human faeces: its identification, purification and characterization. Clin. Sci. *82*, 447–454.
67. Tsai, H.H., Dwarakanath, A.D., Hart, C.A., Milton, J.D., and Rhodes, J.M. (1995). Increased faecal mucin sulphatase activity in ulcerative colitis: a potential target for treatment. Gut *36*, 570–576.
68. Hoffman, J.A., Badger, J.L., Zhang, Y., Huang, S.H., and Kim, K.S. (2000). *Escherichia coli* K1 *aslA* contributes to invasion of brain microvascular endothelial cells in vitro and in vivo. Infect. Immun. *68*, 5062–5067.
69. Pethe, K., Aumercier, M., Fort, E., Gatot, C., Locht, C., and Menozzi, F.D. (2000). Characterization of the heparin-binding site of the mycobacterial heparin-binding hemagglutinin adhesin. J. Biol. Chem. *275*, 14273–14280.
70. Menozzi, F.D., Rouse, J.H., Alavi, M., Laude-Sharp, M., Muller, J., Bischoff, R., Brennan, M.J., and Locht, C. (1996). Identification of a heparin-binding hemagglutinin present in mycobacteria. J. Exp. Med. *184*, 993–1001.
71. Menozzi, F.D., Bischoff, R., Fort, E., Brennan, M.J., and Locht, C. (1998). Molecular characterization of the mycobacterial heparin-binding hemagglutinin, a mycobacterial adhesin. Proc. Natl. Acad. Sci. USA *95*, 12625–12630.
72. Schirmer, A., and Kolter, R. (1998). Computational analysis of bacterial sulfatases and their modifying enzymes. Chem. Biol. *5*, R181–R186.
73. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
74. Higgins, D.G., and Sharp, P.M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene *73*, 237–244.
75. Sonnhammer, E.L.L., and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene *167*, GC1–GC10.
76. Felsenstein, J. (1989). PHYLIP–Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.
77. Sibbald, P.R., and Argos, P. (1990). Weighting aligned protein or nucleic acid sequences to correct for unequal representation. J. Mol. Biol. *216*, 813–818.
78. http://www.ncbi.nlm.nih.gov
79. http://pfam.wustl.edu
80. http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi
81. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.