

Sulfotransferase structural biology and inhibitor discovery

Virginia L. Rath, Dawn Verdugo and Stefan Hemmerich

Sulfotransferases catalyze the transfer of a sulfonyl group from 3'-phosphoadenosine 5'-phosphosulfate (PAPS) to proteins, carbohydrates and small molecules. The sulfotransferases comprise cytosolic and Golgi-resident enzymes; Golgi-resident enzymes represent fertile territory for identifying pharmaceutical targets. Structure-based sequence alignments indicate that the structural fold, and the PAPS-binding site, is conserved between the two classes. Initial efforts to identify sulfotransferase inhibitors by screening kinase inhibitor libraries yielded competitive inhibitors of PAPS with μM IC_{50} values. Within particular classes of Golgi-resident sulfotransferases that show tight *in vitro* specificity, the substrate-binding site might be a suitable drug target, although sulfotransferases are generally assumed to be difficult to inhibit as a result of the expected size and chemical character of the substrate-binding site.

Virginia L. Rath*
Dawn Verdugo
Stefan Hemmerich
Thios Pharmaceuticals
5980 Horton Street
Suite 400
Emeryville
CA 94608, USA
*e-mail:
virginia_rath@yahoo.com

▼ Since the elucidation of the structure of murine estrogen sulfotransferase (mEST) – SULT1E1 – in 1997 [1], the crystal structures of eight more sulfotransferases (STs) have been determined (Table 1). Crystal structures for the phenol ST (SULT1A1 [2]), catecholamine ST (SULT1A3 [3]), dehydroepiandrosterone ST (SULT2A1 [4]) and two isozymes of SULT2B1 (cholesterol ST and pregnenolone ST [5]) have been characterized. In addition, the structure of the ST from *Mycobacterium tuberculosis*, which is responsible for generating the core trehalose-2-sulfate moiety of potential virulence factor sulfolipid-1, has been solved [6]. Crystal structures of STs in complex with 3'-phosphoadenosine 5'-phosphosulfate (PAPS), or the non-sulfated derivative 3'-phosphoadenosine 5'-phosphate (PAP), reveal the conserved nature of the cofactor-binding site, suggesting that all STs have similar catalytic mechanisms of action. The catalytic site of each ST must accommodate diverse substrates, and complexes of the cytosolic STs with substrates have provided information on substrate specificity.

Significantly less is known about the structures of Golgi-resident STs, which are type II single-span membrane proteins with a short N-terminal cytosolic tail and a C-terminal catalytic domain localized to the Golgi lumen. A region of variable length, termed the stem region, is located between the membrane-spanning sequence and the catalytic domain. To date, only two structures of Golgi-resident enzymes are available and both are of heparan sulfate STs. The crystal structure of the ST domain of the dual function heparan sulfate *N*-deacetylase-*N*-ST (NDST-1) complexed with PAP was the first heparan ST structure to be determined [7]. The ST domain for this Golgi-resident enzyme was identified by its similarity to EST sequences, thereby enabling the crystallization of an active fragment containing the ST domain. More recently, the crystal structure of murine heparan sulfate-3-O-ST isoform-1 (3OST-1) complexed with PAP has been solved [8]. 3OST-1 catalyzes the transfer of a sulfonyl moiety to the 3-OH position of a glucosamine residue to form 3-O-sulfoglucosamine, which is the last step in the synthesis of the anticoagulant heparan sulfate. The 3D coordinates of NDST-1 and 3OST-1 are closely related, specifically, all the significant secondary structural elements are conserved between the two enzymes.

There is considerable interest in the structure of the family of Golgi-resident galactose/*N*-acetylgalactosamine/*N*-acetylglucosamine (Gal/GalNAc/GlcNAc)-6-O-STs because of their potential as drug targets. Within this family, the most conclusive evidence for a role in human disease links human *N*-acetylglucosamine-6-O-sulfotransferase (GST)-3 (also known as HEC-GlcNAc6ST, GlcNAc6ST-2, LSST and CHST4) to inflammatory diseases, which is the focus of our first article in the previous issue of *Drug Discovery Today* [9]. Apart from

Table 1. Sulfotransferase structures

Gene	Enzyme	Species	NCBI accession no.	PDB ID	Ligands	Resolution (Å)
Cytosol-localized enzymes						
<i>SULT1A1</i>	Phenol ST	<i>Homo sapiens</i>	NP_001046	1LS6	<i>p</i> -Nitrophenol and PAP	1.90
<i>SULT1A3</i>	Catecholamine ST or dopamine ST	<i>H. sapiens</i>	NP_003157	1CJM	Sulfate	2.40
<i>SULT1E1</i>	Estrogen ST	<i>H. sapiens</i>	NP_005411	1G3M	PAP and 3,5,3',5'-tetra-chloro-biphenyl-4,4'-diol	1.70
<i>Sult1e1</i>	Estrogen ST	<i>Mus musculus</i>	Mutant V269E	1HY3	PAPS	1.80
			NP_075624	1AQY	PAP	1.75
			1AQU	PAP and 17-β estradiol	1.60	
			1BO6	PAP and vanadate	2.10	
<i>SULT2A1</i>	Dehydroepiandrosterone ST	<i>H. sapiens</i>	NP_003158	1EFH	PAP	2.40
				1J99	3-β-hydroxy-5-androsten-17-one	1.99
				1OV4	Aetiocholanolone sulfate	2.70
<i>SULT2B1A</i>	Pregnenolone ST	<i>H. sapiens</i>	NP_004596	1Q1Q	PAP	2.91
<i>SULT2B1B</i>	Cholesterol ST	<i>H. sapiens</i>	NP_814444	1Q1Z	PAP	2.40
				1Q20	PAP and pregnenolone	2.30
				1Q22	PAP and DHEA	2.50
Golgi-resident enzymes						
<i>Hs3st-1</i>	Heparan 3-O-ST isoform-1	<i>M. musculus</i>	NP_034604	1S6T	PAP and sulfate	2.50
<i>NDST-1</i>	Heparan <i>N</i> -deacetylase– <i>N</i> -ST-1	<i>H. sapiens</i>	NP_001534	1NST	PAP	2.30
Bacteria						
<i>Stf0</i>	Stf0 ST	<i>Mycobacterium smegmatis</i>	1TEX_D	1TEX	Trehalose	2.60

Table of sulfotransferases for which crystal structures have been determined to date. Abbreviations: DHEA, 3-β-hydroxy-5-androsten-17-one; NCBI, National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>); PAP, 3'-phosphoadenosine 5'-phosphate; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; PDB ID, Protein Data Bank (<http://www.rcsb.org/pdb>) identification code; ST, sulfotransferase.

two short motifs, there is little sequence similarity between the Gal/GalNAc/GlcNAc STs and the cytosolic enzymes. Despite this, comparison of the sequences with the known crystal structures enabled the identification of a minimal catalytic domain of ~300 residues that is present in the cytosolic- and Golgi-localized enzymes. Analysis of structure-based sequence alignments and secondary structure predictions led to the proposal that the catalytic domains of this family will have the same structural fold as the STs for which structures have already been determined. The majority of the secondary structure elements identified crystallographically can be identified within the Gal/GalNAc/GlcNAc-6-O-ST sequences. Although the exact boundaries of the secondary structure elements are unlikely to be accurately defined from secondary structure predictions, their general location within the primary sequence can be identified with confidence.

The 3'-phosphoadenosine 5'-phosphate-binding site

The crystal structure of mEST complexed with PAP was the first to reveal the details of the cofactor-binding site (Figure 1a) [1]. Comparative analysis showed that the catalytic core of mEST was most closely related to uridylate kinase, adenylate kinase and guanylate kinase. This relationship is the result of the identical connectivity in the nucleotide-binding domains of the four enzymes, including a strand-loop-helix motif termed the phosphate-binding loop, or P-loop. Residues in the P-loop of mEST form the same interactions with the 5'-phosphate of PAP as those formed between the P-loop of uridylate kinase with the β-phosphate of ADP.

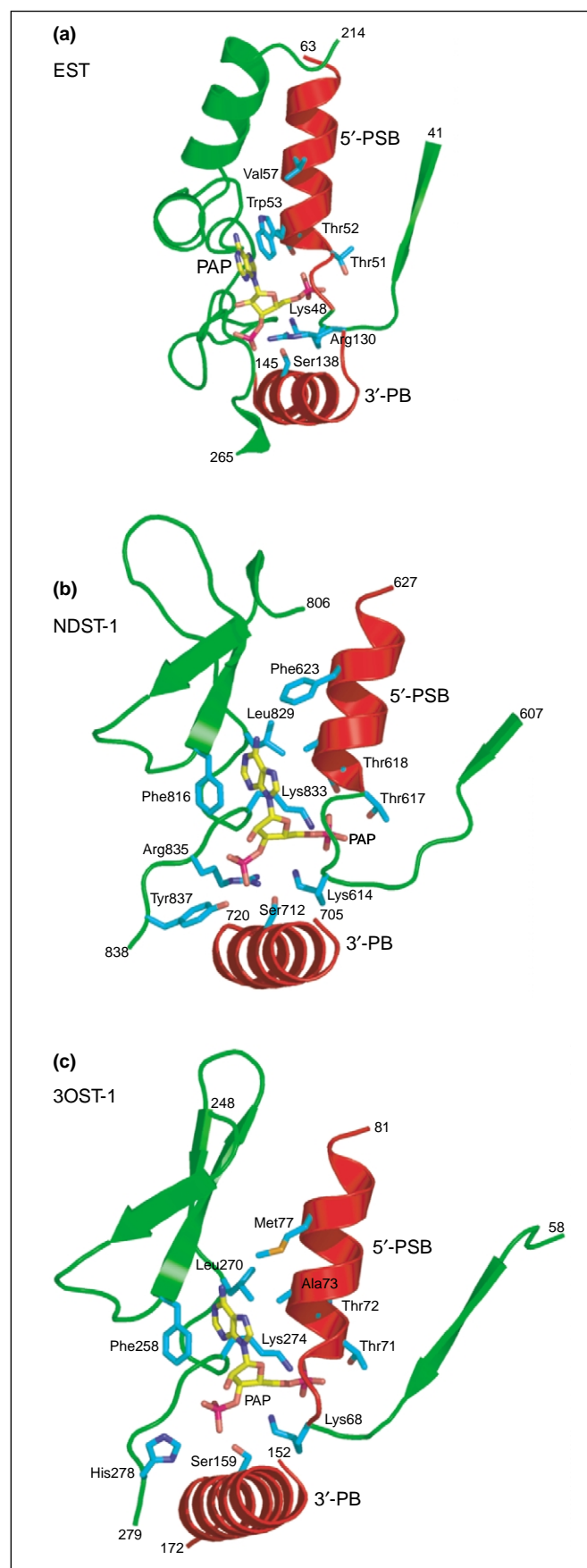
Further sequence analysis revealed two sequence motifs conserved in all STs that are termed the 5'-phosphosulfate-binding (5'-PSB) and the 3'-phosphate-binding (3'-PB) motifs [10]; the 5'-PSB motif contains the P-loop structural

Figure 1. Comparison of 3'-phosphoadenosine 5'-phosphate-binding sites. Ribbon diagrams of the PAP-binding site in (a) EST (1AQU.pdb [9]), (b) heparan sulfate NDST (1NST.pdb [7]) and (c) heparan sulfate-3OST-1 (1VKJ.pdb [8]). Secondary structural elements containing the 3'-PB and 5'-PSB motifs are colored red, other non-conserved structural elements of the PAP-binding sites are shown in green. Nitrogen atoms are depicted in dark blue, oxygen atoms are red, phosphorus atoms are pink, sulfur atoms are orange, carbon atoms of PAP are yellow and all other carbon atoms are cyan. Figure generated using PyMOL (DeLano Scientific; <http://pymol.org>). Abbreviations: EST, human estrogen sulfotransferase; NDST, *N*-deacetylase-*N*-sulfotransferase; 3OST-1, 3-*O*-sulfotransferase isoform-1; PAP, 3'-phosphoadenosine 5'-phosphate; 3'-PB, 3'-phosphate-binding motif; 5'-PSB, 5'-phosphosulfate-binding motif.

motif. Despite the terminology, the 3'-PB and the 5'-PSB binding sites are not mutually exclusive; residues within the 5'-PSB binding site contact the 3'-phosphate of PAP and vice versa. With one exception, the binding mode of PAP to each of the crystallographically determined STs is known and each example has confirmed the structural role of these conserved motifs. In general, the phosphate oxygen atoms of PAP are bound through backbone amide hydrogen bonds and specific hydrogen bonds between the side chains of Ser, Thr, Arg and Lys residues. The adenine ring is stabilized by π -stacking interactions with Trp or Phe side chains and by a hydrogen bond to N6. These interactions are conserved in the cytosolic STs and in the two structures of the Golgi-resident enzymes. Examination of the PAP-binding site of NDST-1, comparison with that of 3OST-1 and analysis of sequence alignments enables prediction of the key features of the PAP-binding site of GST-3.

In human NDST-1, residues that interact with the 3'-phosphate of PAP originate from two separate parts of the polypeptide – those that are part of the conserved 3'-PB motif (residues 700–715) and residues that are part of a random coil (residues 828–842, Figure 1b). Ser712 of the 3'-PB motif forms a hydrogen bond with one of the oxygen atoms of 3'-phosphate. The random coil supplies two backbone amide nitrogen atoms (Gly834 and Arg835), a backbone carbonyl (Ser832) and the side chain of Tyr837 for interactions with two of the oxygen atoms of 3'-phosphate. In addition, the P-loop presents Lys614, which forms a hydrogen bond with the bridging oxygen atom (O3).

Residues interacting with the 5'-phosphate derive from the conserved 5'-PSB motif (PSB-loop residues 614–617 and residues 618–626 from α -helix 1) [7]. The backbone amide nitrogen atoms of residues 614–618 form hydrogen bonds with the oxygen atoms of the 5'-phosphate. In addition, two side chains (Thr617 and Thr618) form hydrogen bonds with the oxygen atoms of the phosphate group. One residue from the random coil, Lys833, forms a side chain hydrogen bond with an oxygen atom of the 5'-phosphate of PAP.



The orientation of the adenine is determined by a hydrogen bond between the N6 of adenine and the backbone oxygen of Trp817 [7]. In addition, the backbone carbonyl oxygen atoms of Leu781 and Arg782 form water-mediated hydrogen bonds to the N6 and N7 nitrogen atoms of adenine. The ring itself is further stabilized by a π -stacking interaction with the side chain of Phe816, a common feature of nucleotide-binding sites.

Outside the conserved PSB loop and 3'-PB, the binding site for PAP in EST differs significantly from that of NDST-1. In effect, the random coil region of NDST-1, which primarily contacts the 3'-phosphate of PAP, is not conserved in EST. The structure of EST diverges from that of NDST-1 for ~50 residues after α -helix 10 (residues 214–265) such that contacts between EST and the 3'-phosphate derive from different parts of the polypeptide chain compared with the location of the residues forming the equivalent interactions in NDST-1. One additional consequence of the structural divergence in this region of the two STs is that different amino acids are recruited for π -stacking interactions with the adenine ring of PAP. In EST, the adenine ring of PAP stacks over Trp52. In NDST-1, Trp52 is replaced by Ala and the π -stacking partner is Phe816, which is located on the opposite side of the ring relative to Trp52 in EST. Furthermore, in NDST-1, to optimize the hydrophobic interactions with Phe816, the adenine ring is rotated ~120 degrees.

The specific substrate recognition of STs for 3'-phosphate results in PAP being preferentially bound over other adenine nucleosides (including AMP, ADP and ATP). Hydrogen bonds are formed between two oxygen atoms of the 3'-phosphate and side chains, and another oxygen atom is linked to main chain atoms by additional hydrogen bonds. The remaining phosphate oxygen is exposed to bulk solvent and does not interact with the protein. The 2'-hydroxyl of the sugar forms a hydrogen bond with a water molecule and is therefore not specifically recognized by the protein. It is probable that recognition of the 2'-hydroxyl is not crucial (as it is in distinguishing DNA from RNA) because of the predominance of interactions between the protein and the 3'-phosphate. If so, this region of the molecule could be exploited to modify inhibitors to improve properties such as solubility or metabolic stability.

Specificity for adenine over other bases is determined by the backbone carbonyl of Trp817, which forms a hydrogen bond with the exocyclic N6 of adenine. The close proximity of the main chain carbonyl would make binding of a guanine (with a carbonyl at this position) unfavorable. The preference for purines over pyrimidines could be attributed to the size of the hydrophobic pocket. Ala619, Phe623, Leu829 and Phe816 pack tightly against the adenine ring

and thus the smaller pyrimidine ring would reduce the extent of these hydrophobic contacts, leading to a weaker interaction.

The PAP-binding site of NDST-1 contains unique features that are not found in crystal structures of cytosolic STs. In particular, the interactions between PAP and residues of the random coil (C-terminal residues 828–842) are unique to NDST-1. The lack of secondary structure in this coil suggests that, in the absence of PAP, this region of the polypeptide might be disordered or in a different conformation. The situation could be different in the full-length NDST-1 protein. The crystal structure contains only the ST domain; residues 1–557 of the intact protein (presumably the N-deacetylase domain) were excluded from the expressed protein. During catalysis, the product of the N-deacetylase reaction is passed to the ST active site and hence it is probable that the missing domain is physically located near the PAP-binding site and potentially has a role in structuring or stabilizing the interactions between PAP and the random coil in intact NDST-1.

The structure of 3OST-1, which was crystallized in the presence of PAP, presents the random coil in the same position relative to the rest of the structure, although there are small differences in the backbone conformation compared with NDST-1 (Figure 1c). The C α carbons of the coil overlap identically in the region of the PAP-binding site but differ in backbone conformation between residues 277–282 (equivalent to 836–841 of NDST-1). Only one side chain (His278) that interacts with PAP is affected by this change in the conformation of the backbone. Although 3OST-1 has a single catalytic domain (NDST-1 has two), the fold is similar to the structure of the ST domain of NDST-1, indicating that the second catalytic domain of NDST-1 is not required to stabilize the conformation of the random coil. Indeed, the amino acid sequence of the random coil is conserved among the heparan STs. This suggests that the random coil interactions with PAP are common to the heparan STs and a feature that distinguishes the heparan PAP-binding site from the cytosolic ST PAP-binding site. If this is the case, it might be possible to design inhibitors that preferentially target one class of STs. It remains unclear whether or not PAP is required to stabilize the conformation of the two classes of ST: the solution of a crystal structure of one of the heparan STs in the absence of PAP will resolve this uncertainty.

The overall fold of 3OST-1 is most similar to that of the ST domain of NDST-1, with a root mean square deviation of 1.3 Å for 230 C α carbon atoms. The sequence identity between these two enzymes is 28%. All the secondary structural elements are conserved, with significant differences found only in the loop regions. The conformation of

PAP bound to 3OST-1 is identical to that found in NDST-1 (but differs from the conformation of PAP bound to EST).

The catalytic site of 3OST-1 is remarkably similar to the ST domain of NDST-1. Although there are five non-conservative amino acid changes in 3OST-1 compared with NDST-1, the structural interactions between the protein and PAP are mostly unchanged. An example of the conservation of contacts is the replacement of Gly69 in 3OST-1 by a Thr in NDST-1. Because the interaction with PAP involves only the main chain nitrogen atom in both proteins, no contacts with PAP are altered. 3OST-1 conserves the van der Waals contact made by Tyr837 in NDST-1 to the N6 nitrogen atom of adenine. The side chain of His278 conserves the hydrogen bond formed between the 3'-phosphate of PAP and the side chain of Tyr837 in NDST-1. Finally, there are two interesting compensatory changes in 3OST-1 involving Arg151, which packs against Ile225. The corresponding interactions in NDST-1 have their amino acid identities reversed; in NDST-1, Ile704 (in the position equivalent to Arg151 of 3OST-1) packs against Arg782 (in the position equivalent to Ile225 of 3OST-1). In 3OST-1, Arg151 forms a hydrogen bond with the 3'-phosphate of PAP. Although Arg782 (NDST-1) is oriented away from PAP, it could presumably move to form the same hydrogen bond. The single exception to this conservation of contacts is a Thr to Arg substitution; the side chain hydrogen bond between Thr618 and the 5'-phosphate cannot be made by Arg72 of 3OST-1.

Chemical tractability of the 3'-phosphoadenosine 5'-phosphate-binding site

Will it be possible to create potent and specific ST inhibitors that are selective for the PAP-binding site? Indeed, inhibitors have been developed that exploit the related purine-binding site occupied by ATP in kinases (e.g. Glivec [11]). The challenge is to create ATP or, in the case of GST-3, PAP analogs that lack phosphate groups. Phosphate-containing molecules are typically unstable and lack bioavailability. A general approach to compensate for the lack of phosphates is to add moieties that simultaneously increase hydrophobic interactions and retain the hydrogen bonding capability of the adenine ring. Often, hydrophobic pockets (in the order of 100 Å³ or more in volume) near the adenosine-binding site can be exploited to achieve these objectives.

Are there suitable cavities in NDST-1 or 3OST-1? Although there are no hydrophobic cavities in the adenosine-binding site of either of these STs, the presence of water molecules in their respective active sites suggests that the docking of PAP in the active site is not optimal; 50% of the total surface area of PAP is solvent accessible. In NDST-1, the N6

and N7 nitrogen atoms of adenine form hydrogen bonds with water molecules. A common medicinal chemistry strategy in such a case is to integrate the water molecules directly into a suitable chemical scaffold. Alternatively, the space occupied by the water molecules could be replaced by substituents on the N7 or N6 to extend the ligand into a region of hydrophobic residues bordered by Ala619, Phe623, Leu781, Arg782 (aliphatic backbone), Pro785, Leu808, Trp817 and Cys818. A similar scenario can be written for 3OST-1, where the positions of the water molecules are conserved compared with NDST-1.

Homology modeling of GST-type sulfotransferases

Are the cytosolic or heparan STs good homology models for the Gal/GalNAc/GlcNAc-6-O-STs? Apart from the PAP 3'-PB and 5'-PSB motifs, there is limited sequence similarity between the Gal/GalNAc/GlcNAc-6-O-STs and the other STs. Structure-based sequence alignments and secondary structure predictions of members of the GST family strongly suggest that the overall fold of each member is likely to be the same as the other STs. Within that context, it is to be anticipated that the architectural details of the catalytic site will be sufficiently different that a crystal structure of the specific GST target will be required for structure-based drug design. Differences in the carbohydrate substrates used by each enzyme support this hypothesis.

Comparison of ST sequences revealed that the most significant difference between GST-3 and either of the heparan STs is the presence of a large insertion in the primary sequence. GST-3 contains a 68 amino acid insertion (residues 84–152), which has been mapped to a location on NDST-1 between the second (630–632) and third (672–676) strands of the β -sheet of the Rossmann fold (unpublished results). This inserted region contains six of the eight cysteine residues within the ST domain, four of which are predicted to form intramolecular disulfide bonds. The positioning of this insert within the primary sequence locates it in 3D space close to the region occupied by the N-terminus (near residue 60 of NDST-1) in the heparan STs.

Modeling of the cofactor-binding site of GST-3

If the equivalent residues of GST-3 from the conserved 3'-PB and 5'-PSB motifs are mapped onto the crystal structure of NDST-1, the resulting model suggests that many of the interactions between GST-3 and PAP will be conserved. Eleven residues of the cofactor-binding site are either identical to those from the binding site of NDST-1 or involve conservative changes that are not expected to alter the contacts to PAP. In addition, there are nine non-conservative substitutions; of these, five involve main chain interactions

(i.e. side chain differences should have no functional consequences). In NDST-1, Lys833 corresponds to Asn323 in GST-3 and it is probable that the side chain of Asn could form the same hydrogen bond as the lysine side chain. By contrast, the side chain of Tyr837 (NDST-1) forms a hydrogen bond with the 3'-phosphate, which cannot be formed in GST-3 where the corresponding residue is Ala327.

The adenine ring in all the ST structures is supported by a stacking interaction with Phe, Trp or Tyr. In NDST-1, the adenine ring is sandwiched between Phe816 and Ala619. In GST-3, the Phe is replaced by Lys313 and the Ala is replaced by Phe57, thus the side chains are switched but the extent of the hydrophobic contacts with the adenine ring will probably be preserved. Likewise, the hydrophobic contacts formed by Phe623, which packs orthogonally (edge of Phe to the face of adenine) to the adenine ring, could be replaced by hydrophobic interactions to Leu61 in GST-3. One architectural difference in the binding site of GST-3 compared with that of NDST-1 is the absence of the Cys818–Cys828 disulfide bond; in GST-3, these residues are replaced by Met315 and His318, respectively. The intramolecular disulfide bond in NDST-1 stabilizes a ten residue β extension that is absent in GST-3, where the β extension is replaced by a simple turn.

Mutational analysis of EST [12], flavonol 3-ST [13] and the crystal structure of EST with vanadate (a transition state analog [12]) strongly implicated residues Lys48 and His108 in the catalytic mechanism of STs. Kakuta *et al.* [12] proposed that Lys48 stabilized the transition state and His108 acted as a catalytic base and/or participated in transition state stabilization. Lys48 of EST is conserved in NDST-1 (Lys614), but is replaced by an Arg residue (Arg52) in GST-3. Interestingly, the Lys to Arg substitution in EST retained low activity, which suggests that Arg52 could stabilize the transition state in GST-3. His108 is not conserved in either heparan ST and, based on the sequence alignment, is not conserved in GST-3, where the equivalent amino acid is Val (Val179).

The substrate-binding site

In contrast to the highly conserved nature of the PAP-binding site, the substrate-binding site of each ST reflects differences in specificity. The cytosolic STs typically recognize small hydrophobic molecules such as steroids, phenols and catecholamines, whereas the Golgi-resident STs recognize hydrophilic carbohydrates or Tyr residues in peptides. Complexes of cytosolic estrogen, phenol, dehydroepiandrosterone and cholesterol STs with substrates have been determined (Table 1). No crystal structure of a carbohydrate ST complexed with its substrate has been solved, although the

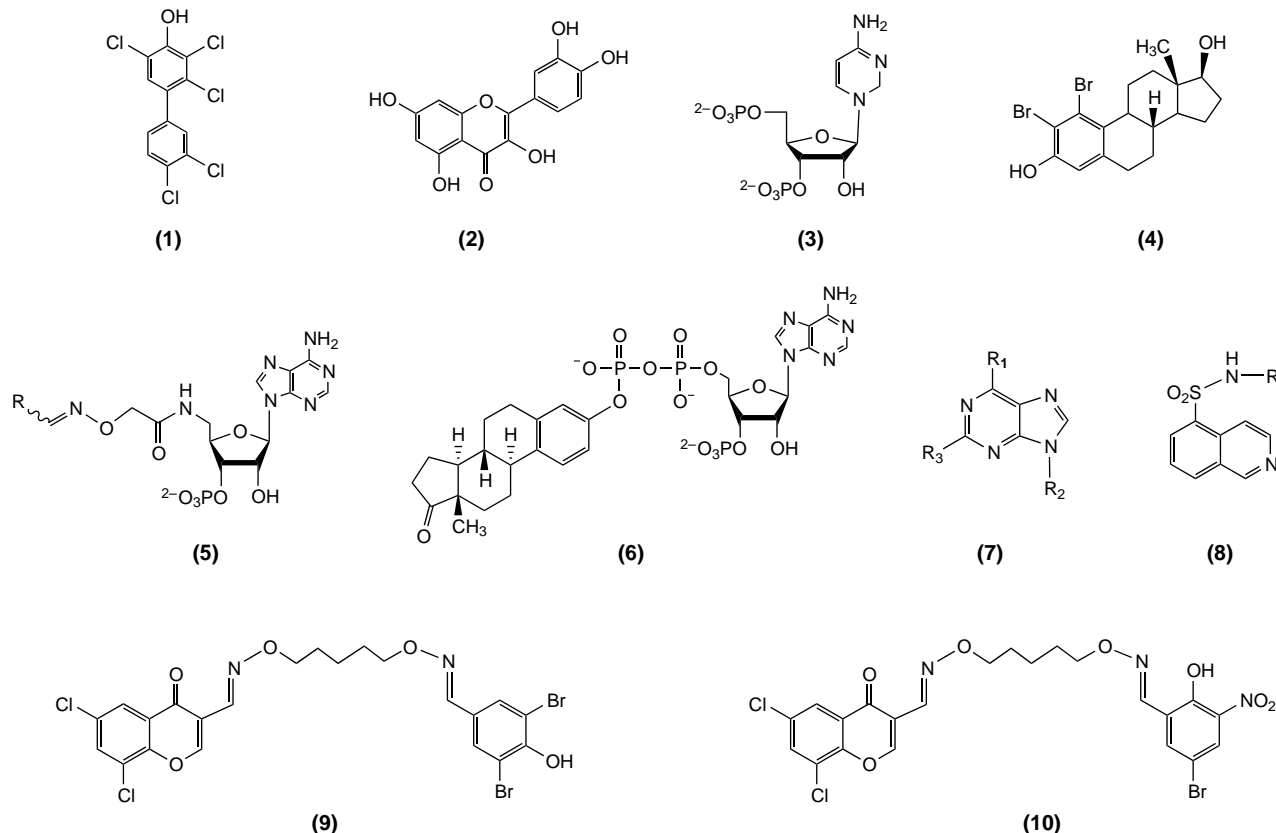
binding sites have been inferred from the known structures. To date, targeting the substrate-binding site or developing bisubstrate inhibitors of the Golgi-resident STs has not been attempted. Because the carbohydrate STs exhibit unique substrate preferences, this approach could have particular value in distinguishing between the members of the GST family, which share 30–50% sequence identity, as well as between the carbohydrate and cytosolic STs.

Crystallographic studies have shown that sugars typically bind to hydrophilic grooves on the surfaces of proteins with affinities in the mM to μ M range [14]; this observation is usually interpreted as an indication that carbohydrate-binding sites are not 'druggable'. There are also examples of small saccharides buried in solvent inaccessible clefts within a protein; in these cases, the affinities range from μ M to nM. The substrate-binding site of NDST-1 is thought to consist of residues that line the open cleft ($12 \times 8 \times 8$ Å in size) perpendicular to the PAP-binding site [7], which is consistent with the description of a large hydrophilic groove. The recognized substrate for NDST-1 is a heparan polymer, which necessitates the large size of the binding site. Residues of α -helix 6 (706–719) and residues within the random coil (633–648, just after strand β 2) are potential candidates for substrate specific interactions.

All of the GlcNAc-6-O-ST family members (GST-2, GST-3, GST-4 α , GST-4 β and GST-5) modify the 6-hydroxyl group of a terminal GlcNAc [15], a requirement that suggests the binding site of these enzymes is smaller and more druggable than that found in NDST-1 or 3OST-1. *In vitro* studies have shown that a monosaccharide appended with a chemical handle (e.g. β -benzyl-GlcNAc) is a substrate for GST-3 [16]. Although disaccharides are accommodated to the same extent as monosaccharides, trisaccharides are not [17]; this suggests that the catalytic site of GST-3 could be inhibited by a compound with a molecular weight suitable for drug development. GST-2 and GST-3 exhibit distinct preferences for functional groups on disaccharide substrates, suggesting that enzyme-specific glycomimetics could be designed [18]. The ensuing challenge would be to convert such a glycomimetic into a metabolically stable drug.

Design and discovery of sulfotransferase inhibitors

To date, efforts to develop inhibitors for the STs have focused on exploring the features of sulfate transfer, rather than drug development. Hence, molecules such as environmental toxins, natural products and bisubstrate-based compounds [i.e. compounds that incorporate elements from the substrate and the cofactor (PAPS)] have been studied, primarily in the context of the cytosolic STs. Polychlorinated biphenols (1, Figure 2 [19]), dietary agents such as



Drug Discovery Today

Figure 2. Examples of sulfotransferase inhibitors. Compounds 1–6 are inhibitors of EST. Compound 1, an environmental toxin, is a hydroxylated polychlorinated biphenol; compound 2 is quercetin, a flavonoid; compounds 3 and 4 are inhibitors that mimic products of the ST reaction; compound 5, a 3'-phosphoadenosine inhibitor; compound 6 (IC_{50} of 10 nM) is designed to mimic elements of PAPS and substrate – 6 was found to be competitive with PAP but not substrate; compound 7 is the core structure of a series of compounds that inhibit EST and/or NodH, identified from a focused library designed to target the kinase CDK2; compound 8 is the core structure of the isoquinoline sulfonamide kinase inhibitors, which are also active against EST, NodH, GST-2 and GST-3 (Table 2); and compounds 9 (IC_{50} of 30 μ M) and 10 (IC_{50} of 40 μ M) are TPST-2 inhibitors. Structures generated using ChemDraw Ultra 7.0 (CambridgeSoft; <http://www.cambridgesoft.com>). Abbreviations: CDK2, cyclin-dependent kinase 2; EST, human estrogen sulfotransferase; GST, human *N*-acetylglucosamine-6-O-sulfotransferase; NodH, prokaryotic root nodulation factor ST; PAP, 3'-phosphoadenosine 5'-phosphate; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; ST, sulfotransferase; TPST-2, tyrosyl protein sulfotransferase-2.

quercetin (2, Figure 2 [20]) and compounds that mimic the end products of sulfate transfer (3 and 4, Figure 2 [21]) have been investigated as inhibitors of EST, which is one of the most widely investigated cytosolic STs. Two bisubstrate-based approaches to identifying EST inhibitors have been described. Screening of a 447 member 3'-phosphoadenosine library identified several compounds showing greater than 80% inhibition of EST at 200 μ M (5, Figure 2 [22]). Another compound that incorporated components of donor and acceptor substrates inhibited EST with an IC_{50} of 10 nM (6, Figure 2 [23]). This compound displayed competitive inhibition with respect to PAPS but not substrate, which suggests that the molecule was not binding simultaneously to the PAPS- and substrate-binding sites, possibly because the phosphate linker was too short to span the two sites.

By contrast, there are few examples of Golgi-resident ST inhibitors, primarily as a result of the inherent difficulties in developing robust expression systems and assay methods for these enzymes. One approach towards inhibiting these STs exploits the similarities between the reactions catalyzed by STs and kinases. STs and kinases use adenosine-based donor nucleotides (PAPS for STs and ATP for kinases) to transfer an anionic moiety onto their respective substrates; hence, it was reasoned that ATP derivatives might also function as ST inhibitors. Inhibitors that are active against EST (IC_{50} values of 0.5–35.0 μ M, [24]) and the prokaryotic root nodulation factor ST, NodH [25] (7, Figure 2 [26]), were identified from screens of a 2,6,9-purine derivative library designed to target cyclin-dependent kinase 2 (CDK2) [27], supporting the validity of this approach.

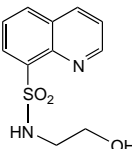
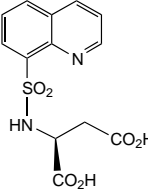
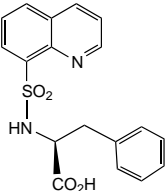
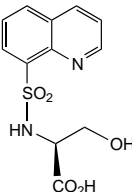
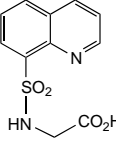
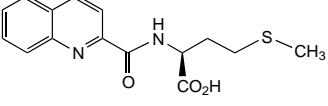
One class of kinase inhibitor, the isoquinoline sulfonamides (8, Figure 2), was of particular interest because of the chemical tractability of the series and the solution of a crystal structure of a complex between cAMP-dependent protein kinase and an isoquinoline showed that the isoquinoline moiety was bound in the subsite occupied by the adenine ring of ATP [28]. The possibility that this class might also be active against the Golgi-resident STs was evaluated by synthesizing a compound library of 100 isoquinoline and quinoline derivatives that were screened against a panel of STs consisting of EST, NodH, GST-2 and GST-3. Some of the most active compounds inhibited a single enzyme with IC_{50} values in the 30 to 100 μ M range with no activity against the other three STs (Table 2 [29]). These preliminary results warrant further investigation of adenine-mimicking heterocyclic libraries, with the goal of increasing the potency and specificity of Golgi-resident ST inhibition.

An alternative strategy taken to identify ST inhibitors involved assembling a library of ligands of diverse molecular weights that carry a common chemical handle. Ligands with inhibitory activity were linked via the chemical handle to form dimers, which were then retested for improved potency. Using this method of ligand assembly, two dimers with activity against the Golgi-resident tyrosyl protein ST-2 (TPST-2), which transfers sulfate to tyrosine side chains in proteins, were identified [30]. The compounds (9 and 10, Figure 2), with IC_{50} values of 30 and 40 μ M, were the first known inhibitors of TPST-2 activity.

Conclusions

Efforts to develop potential therapeutic inhibitors of STs have been spurred by an increasing volume of evidence that validates a link between the Golgi-resident ST, GST-3, and inflammatory diseases. The crystal structures of the Golgi-localized heparan STs and the cytosolic STs are not suitable homology models for GST-3 for the purposes of

Table 2. Selectivity profile of inhibitors tested against a panel of sulfotransferases

Compound	IC_{50} (μ M)			
	EST	NodH	GST-2	GST-3
	>200	50	>200	>200
	100	>200	>200	>200
	>200	>200	75	>200
	>200	>200	100	>200
	>200	>200	>200	100
	>200	>200	>200	30

Selectivity profile of leads from a library of isoquinoline and quinoline derivatives tested against a panel of sulfotransferases [28]. Some of the most active compounds inhibit a single enzyme with IC_{50} values in the 30–100 μ M range. >200 indicates no inhibition at 200 μ M, the highest concentration tested. Structures created using ChemDraw Ultra 7.0 (CambridgeSoft; <http://www.cambridgesoft.com>). Abbreviations: EST, human estrogen sulfotransferase; GST, *N*-acetylglucosamine-6-*O*-sulfotransferase; NodH, prokaryotic root nodulation factor ST.

drug design. However, structure-based sequence alignments of the cytosolic STs and Golgi-resident STs suggest that the overall fold of the two classes will be the same. Conservation of the fold and two sequence motifs involved in binding PAPS suggests that many of the interactions between the enzyme and the co-substrate will be

found in all STs. Past efforts in the development of inhibitors for cytosolic STs and Golgi-resident STs have focused on understanding the mechanism of enzyme action more than drug development. The growing interest in GST-3 has fueled renewed efforts to exploit the similarity between PAPS and ATP to identify PAPS-competitive inhibitors from focused libraries designed to inhibit kinases. Compounds with μM IC_{50} values were identified, reinforcing the applicability of the approach; medicinal chemistry efforts will be required to develop these molecules into drugs. These efforts have used small, designed libraries and the screening of a large pharmaceutical library has yet to be attempted. Thus, only a fraction of the possible chemical space has been searched, and interesting results can be expected as work progresses.

Acknowledgements

We acknowledge Andrew May for his efforts in sequence and structural analysis of the STs. We are grateful to Stephen Harrison and Steven D. Rosen for their appraisal of the manuscript.

References

- Kakuta, Y. *et al.* (1997) Crystal structure of estrogen sulfotransferase. *Nat. Struct. Biol.* 4, 904–908
- Gamage, N. *et al.* (2003) Structure of a human carcinogen-converting enzyme, SULT1A1. *J. Biol. Chem.* 278, 7655–7662
- Bidwell, L. *et al.* (1999) Crystal structure of human catecholamine sulfotransferase. *J. Mol. Chem.* 293, 521–530
- Rehse, P. *et al.* (2002) Crystal structure of human dehydroepiandrosterone sulfotransferase in complex with substrate. *Biochem. J.* 364, 165–171
- Lee, K. *et al.* (2003) Crystal structure of human cholesterol sulfotransferase (SULT2B1b) in the presence of pregnenolone and 3'-phosphoadenosine 5'-phosphate: rationale for specificity differences between prototypical SULT2A1 and the SULT2B1 isoforms. *J. Biol. Chem.* 278, 44593
- Mougous, J. *et al.* (2004) Identification, function and structure of the sulfotransferase that initiates sulfolipid-1 biosynthesis in *Mycobacterium tuberculosis*. *Nat. Struct. Biol.* 8, 721–729
- Kakuta, Y. *et al.* (1999) Crystal structure of the sulfotransferase domain of human heparan sulfate N-deacetylase/N-sulfotransferase 1. *J. Biol. Chem.* 274, 10673–10676
- Edavettal, S.C. *et al.* (2004) Crystal structure and mutational analysis of heparan sulfate 3-O-sulfotransferase isoform 1. *J. Biol. Chem.* 279, 25789–25797
- Hemmerich, S. *et al.* (2004) Strategies for drug discovery targeting sulfation pathways. *Drug Discov. Today* 9, 967–975
- Kakuta, Y. *et al.* (1998) Conserved structural motifs in the sulfotransferase family. *Trends Biochem. Sci.* 23, 129–130
- Capdeville, R. *et al.* (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* 1, 493–502
- Kakuta, Y. *et al.* (1998) The sulfuryl transfer mechanism. Crystal structure of a vanadate complex of estrogen sulfotransferase and mutational analysis. *J. Biol. Chem.* 273, 27325–27330
- Marsolais, F. and Varin, L. (1995) Identification of amino acid residues critical for catalysis and cosubstrate binding in the flavonol 3-sulfotransferase. *J. Biol. Chem.* 270, 30458–30463
- Weis, W.L. and Drickamer, K. (1996) Structural basis of lectin carbohydrate recognition. *Annu. Rev. Biochem.* 65, 441–473
- Bowman, K.G. *et al.* (2001) Biosynthesis of L-selectin ligands: sulfation of sialyl Lewis x-related oligosaccharides by a family of GlcNAc-6-sulfotransferases. *Biochemistry* 40, 5382–5391
- Bhakta, S. *et al.* (2000) Sulfation of N-acetylglucosamine by chondroitin 6-sulfotransferase 2 (GST-5). *J. Biol. Chem.* 275, 40226–40234
- Bowman, K.G. *et al.* (1998) Identification of an N-acetylglucosamine-6-O-sulfotransferase activity specific to lymphoid tissue: an enzyme with a possible role in lymphocyte homing. *Chem. Biol.* 5, 447–460
- Cook, B.N. *et al.* (2000) Differential carbohydrate recognition of two GlcNAc-6-sulfotransferases with possible roles in L-selectin ligand biosynthesis. *J. Am. Chem. Soc.* 122, 8612–8622
- Kester, M.H.A. *et al.* (2000) Potent inhibition of estrogen sulfotransferase by hydroxylated PCB metabolites: a novel pathway explaining the estrogenic activity of PCBs. *Endocrinology* 141, 1897–1900
- Rozhin, J. *et al.* (1977) Studies on bovine adrenal estrogen sulfotransferase. Inhibition and possible involvement of adenine-estrogen stacking. *J. Biol. Chem.* 252, 7214–7220
- Horwitz, J.P. *et al.* (1978) IV. Synthesis and assay of analogs of adenosine 3',5'-diphosphate as inhibitors of bovine adrenal estrogen sulfotransferase. *Biochim. Biophys. Acta* 525, 364–372
- Armstrong, J.I. *et al.* (2001) A library approach to the generation of bisubstrate analogue sulfotransferase inhibitors. *Org. Lett.* 3, 2657–2660
- Armstrong, J.I. *et al.* (2003) Synthesis of a bisubstrate analogue targeting estrogen sulfotransferase. *J. Org. Chem.* 68, 170–173
- Verdugo, D.E. *et al.* (2001) Discovery of estrogen sulfotransferase inhibitors from a purine library screen. *J. Med. Chem.* 44, 2683–2686
- Ehrhardt, D.W. *et al.* (1995) *In vitro* sulfotransferase activity of NodH, a nodulation protein of *Rhizobium meliloti* required for host-specific nodulation. *J. Bacteriol.* 177, 6237–6245
- Armstrong, J.I. *et al.* (2000) Discovery of carbohydrate sulfotransferase inhibitors from a kinase-directed library. *Angew. Chem. Int. Ed. Engl.* 39, 1303–1306
- Chang, Y.T. *et al.* (1999) Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem. Biol.* 6, 361–375
- Engh, R.A.G. *et al.* (1996) Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89 - structural implications for selectivity. *J. Biol. Chem.* 271, 26157–26164
- Verdugo, D.E. *et al.* (2003) Small molecule inhibitors of the sulfotransferases. In *Carbohydrate-based Drug Discovery* (Vol. 2) (Wong, C.H., ed.), pp. 781–802, Wiley-VCH
- Kehoe, J.W. *et al.* (2002) Tyrosylprotein sulfotransferase inhibitors generated by combinatorial target-guided ligand assembly. *Bioorg. Med. Chem. Lett.* 12, 329–332