From Steroid and Drug Metabolism to Glycobiology, Using Sulfotransferase Structures to Understand and Tailor Function

By Lars C. Pedersen¹*, MyeongJin Yi², Lee G. Pedersen¹,³, Andrea M. Kaminski²

¹Genome Integrity and Structural Biology Laboratory or ²Reproductive and Developmental Biology Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, 111 TW Alexander Dr., Bldg. 101/Rm F338, Research Triangle Park, North Carolina, USA, 27709.

³Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, 27599.

*Corresponding Author (email: pederse2@niehs.nih.gov)

This work was supported in part by Division of Intramural Research of the National Institute of Environmental Health Sciences, National Institutes of Health Grants 1ZIA ES102645 (to L.C.P.).
Running Title: Structural Biology of Sulfotransferases

Number of pages: 66

Number of tables: 2

Number of figures: 8

Number of references: 188

Number of words in Abstract: 193

Number of words in Body: 8,146

Abbreviations: APS, adenosine 5'-phosphosulfate; GAG, glycosaminoglycans; GlcA, glucuronic acid; GlcNAc, N-acetylglucosamine; HSSTs, heparan sulfate sulfotransferases; IdoA, Iduronic acid; MD, molecular dynamics; NDST-1, N-deacetylase/N-sulfotransferase isoform 1; NMR, nuclear magnetic resonance; NST-1, sulfotransferase domain of NDST-1; OST, O-sulfotransferase; PAPS, 3’-phosphoadenosine 5’-phosphosulfate; PAP, 3’-phosphoadenosine 5’-phosphate; PDB, protein data bank; PSB, phosphosulfate binding; SB, sulfate binding; SULTs, cytosolic sulfotransferases; QM/MM, quantum mechanics/molecular mechanics; TPSTs, Tyrosyl protein sulfotransferases;
Abstract:

Sulfotransferases are ubiquitous enzymes that transfer a sulfo group (SO$_3^-$) from the universal cofactor donor PAPS to a broad range of acceptor substrates. In humans, the cytosolic sulfotransferases (SULTs) are involved in the sulfation of endogenous compounds such as steroids, neurotransmitters, hormones and bile acids as well as xenobiotics including drugs, toxins and environmental chemicals. The Golgi associated membrane-bound sulfotransferases are involved in post-translational modification of macromolecules from glycosaminoglycans to proteins. The sulfation of small molecules can have profound biological effects on the functionality of the acceptor, including activation, deactivation or enhanced metabolism and elimination. Sulfation of macromolecules has been shown to regulate a number of physiological and pathophysiological pathways by enhancing binding affinity to regulatory proteins or binding partners. Over the last 25 years, crystal structures of these enzymes have provided a wealth of information on the mechanisms of this process and the specificity of these enzymes. This review will focus on the general commonalities of the sulfotransferases, from enzyme structure to catalytic mechanism as well as providing examples into how structural information is being utilized to either design drugs that inhibit sulfotransferases or to modify the enzymes to improve drug synthesis.

Significance Statement:

This manuscript honors Dr. Masahiko Negishi’s contribution to our understanding of sulfotransferases mechanism, specificity and their roles in biology by analyzing the crystal structures that have been solved over the last 25 years.

Introduction:
The sulfated products of reactions catalyzed by the sulfotransferase family of enzymes are involved in many physiological and pathophysiological processes (Bishop et al., 2007; Coughtrie, 2016; Duffel, 2010 updated 2016; Gamage et al., 2006; Langford et al., 2017; Lindahl et al., 2015). Vertebrate sulfotransferases utilize 3’-phosphoadenosine 5’-phosphosulfate (PAPS) as the universal sulfo donor. PAPS is synthesized by one of two PAPS synthases, which are bifunctional enzymes that utilize two ATP molecules to produce one molecule of PAPS (Figure 1) (Lyle et al., 1994; Xu et al., 2000). The ATP sulfurylase domain displaces pyrophosphate from ATP with inorganic sulfate to produce adenosine 5’-phosphosulfate (APS). APS is subsequently phosphorylated at the 3’ hydroxyl by the APS kinase domain to produce PAPS.

Sulfation by cytosolic sulfotransferases (SULTs) of small molecules on hydroxyl or amine moieties is one of the major pathways for the detoxification of xenobiotics and elimination of endogenous small molecules such as hormones, neurotransmitters and bile acids from the body (Coughtrie, 2016; Duffel, 2010 updated 2016; Gamage et al., 2006). Sulfation of these compounds typically improves water solubility, enhancing elimination. However, sulfation can also result in biologically active compounds. Dehydroepiandrosterone and estrone sulfate can act as circulating intermediates for the biosynthesis of hormones, while pregnenolone sulfate can regulate neurotransmitter receptors (Mueller et al., 2015). As such, sulfation/desulfation is a critical process that modulates steroidogenesis and hormone action in various tissues (Mueller et al., 2015). Inhibition of sulfotransferases due to environmental exposures to chemicals such as polychlorinated biphenols and flame retardants may result in disruption of proper endocrine homeostasis (Hamers et al., 2008; Kester et al., 2000; Kester et al., 2002).
In addition to aiding in the elimination of dietary compounds and environmental toxins, SULTs are major Phase II drug metabolizing enzymes, not only involved in their elimination, but can also activate prodrugs. Sulfation of minoxidil, for example, is critical for its hair growth effects, while sulfation of oxamniquine activates the drug to treat schistosomiasis (Buhl et al., 1990; Pica-Mattoccia et al., 2006; Valentim et al., 2013).

In humans, there are 13 SULT genes encoding 14 proteins, with two of the proteins resulting from alternative splicing (Coughtrie, 2016). These proteins are between 284 to 365 amino acids in length. Historical nomenclature for the SULTs was based upon the original substrate with which they were identified (e.g. EST estrogen sulfotransferase, PST phenol sulfotransferase, and AST aryl sulfotransferase), but as it became clear that many of these enzymes catalyzed substrates with different functional groups and exhibited overlapping specificity, a more systematic nomenclature was needed (Blanchard et al., 2004; Duffel, 2010 updated 2016). In brief, in the current nomenclature, the first numeral after the SULT defines the family (45% sequence identity) the following capital Arabic letter defines the subfamily (60%), with the final numeral defining the isoform. Splice variants are then assigned a lowercase letter (ex. SULT2B1b).

Tyrosyl protein sulfotransferases (TPSTs) are Type-II integral membrane proteins localized in the trans-Golgi network that are responsible for sulfation of the tyrosine side chain of acceptor proteins, representing one of the major post-translational modifications (Figure 1) (Moore, 2003). Tyrosine sulfation can have a number of biological consequences including effecting circulation half-life, proteolytic processing of bioactive peptides, and protein-protein interactions impacting many processes including blood coagulation, inflammation and viral infection (Farzan et al., 1999; Leyte et al., 1991; Moore, 2003; Pouyani and Seed, 1995).
Humans contain two TPSTs (TPST-1 and -2) that share 64% sequence identity and are 370 and 377 amino acids in length, respectively (Beisswanger et al., 1998; Niehrs and Huttner, 1990; Ouyang et al., 1998). These enzymes have broad and slightly different substrate specificities (Mishiro et al., 2006).

There are at least 37 Golgi-associated sulfotransferases in humans, a large number of which are involved in the sulfation of the glycosaminoglycans (GAGs)—heparan sulfate, chondroitin sulfate and dermatan sulfate (Langford et al., 2017; Zerbino et al., 2018). These GAGs are found attached to a select group of specific proteins or lipids and are found ubiquitously on the cell surface and in the extracellular matrix (Bishop et al., 2007; Iozzo and Schaefer, 2015; Lindahl et al., 2015; Lindahl and Li, 2009). As such, they play important roles in how cells communicate with their surrounding environment through interactions with protein binding partners. These GAGs are involved in a long list of physiological and pathophysiological processes including embryonic development, blood coagulation, inflammation, bacterial and viral infection, neurodegenerative diseases, and cancer (Bishop et al., 2007; Lindahl and Li, 2009; Shi et al., 2021). GAGs are made up of linear repeating disaccharide units that can be modified by deacetylation, epimerization, and sulfation. This review will focus on heparan sulfate sulfotransferases due to the availability of structural information. The sulfation of specific hydroxyls and amines on the GAGs is carried out by unique sulfotransferases and appears to be a somewhat ordered process (Multhaupt and Couchman, 2012). In general, heparan sulfates (HS) are synthesized through chain elongation by a set of glycosyltransferases (exotosin 1 and 2) onto a common linker region to produce repeating disaccharides units of N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA). The bifunctional N-deacetylase/N-sulfotransferase (4 isoforms in humans) can deacetylate and
subsequently sulfate the glucosamine at the 2-amine position. This becomes a good substrate for the Cs-epimerase, which can convert adjacent GlcAs into iduronic acids (IdoA). Although the 2-\textit{O}-sulfotransererase (1 isoform) can sulfate the 2-hydroxyl on GlcA, sulfation on IdoA appears to be the preferred substrate and is much more prevalent in biology (Bethea et al., 2008; Liu et al., 2014; Rong et al., 2001). Sulfation by both the 6-\textit{O}-sulfotransferases (3 isoforms) and the 3-\textit{O}-sulfotransferases (7 isoforms) occurs at the 6- and 3- hydroxyl groups of glucosamines, respectively. While some sulfotransferase isoforms appear to have similar substrate specificities, others are very distinct. The extent and types of modification in a given tissue or cell type are spatially and temporally regulated and are determined by which isoforms are present and expressed.

**Historical Prospective**

In 1876, the relevance of sulfotransferases in medicine was first reported when E. Baumann determined that the conjugate of carbolic acid, a common antiseptic used in surgery at the time, could be absorbed by the skin and secreted in the urine as phenol sulphuric ester (Baumann, 1876; Folin and Denis, 1915). Recognizing the importance and relevance of sulfation in chondroitin sulfate, heparin, steroids and phenol detoxification, Robbins and Lipmann reported in 1956 that the biological sulfo donor was indeed PAPS, which could be produced by a liver enzyme with the addition of ATP (Lipmann, 1958; Robbins and Lipmann, 1956). They referred to the enzymes that must utilize PAPS for the sulfo donor as sulfokinases. It was around this same time period that the first tyrosine-\textit{O}-sulfate was found in a peptide from fibrinogen (Bettelheim, 1954). By this time, HS had already been identified as an effective anticoagulant and had been introduced as a widely available therapeutic (Howell and Holt, 1918;
Lim, 2017). It wasn’t until 1987 that the first cytosolic sulfotransferase was unknowingly cloned as an androgen-repressible rat liver protein (Chatterjee et al., 1987). The cloning of the first Golgi sulfotransferases involved in HS biosynthesis occurred in 1992, and the TPSTs followed in 1998 (Hashimoto et al., 1992; Ouyang et al., 1998).

Prior to the first crystal structures, sequence alignments of the SULTs were used to identify critical residues in the functionality of these enzymes (Khan et al., 1993; Roche et al., 1991; Weinshilboum and Otterness, 1994). These alignments revealed two highly conserved regions including an N-terminal region YPKSGTXW and a carboxy terminus GXXGXXK reminiscent of the GXXXXGK[TS] P-loop structures found to bind ATP in kinases (Saraste et al., 1990). In 1997, the Negishi lab at the National Institute of Environmental Health Sciences at the National Institute of Health, published the first crystal structure of a sulfotransferase (mouse estrogen sulfotransferase or mSULT1E1) ushering in a new era of understanding of the molecular function of these enzymes (Kakuta et al., 1997). This structure revealed that both regions indeed bound to the PAPS, however it was the amino terminal region that was structurally related to the P-loop and cradled the 5’ phosphate while the C-terminal region bound to the 3’ phosphate unique to PAPS (Figure 2). The structure also revealed a conservation of a central scaffold as well as substrate positioning similar to that of the adenylate kinase family of enzymes, suggesting a common mechanism between sulfation and phosphorylation. This analogy can be extended to desulfation, since sulfatases display conservation in structure and activity to that of alkaline phosphatase (Bond et al., 1997; O’Brien and Herschlag, 1998). Crystal structures currently exist for 12 of the 14 human SULTs, representatives from each of the vertebrate HS sulfotransferases as well as both TPSTs (Table 1).
In this review, we will delve into what has been learned from crystal structures of the various sulfotransferase families (SULTs, HSSTs and TPSTs) regarding conservation of structure and mechanism, determinants of specificity and how this information is being utilized to design and engineer new pharmacological tools.

**Key Advances in understanding:**

**Conservation of structure and mechanism among sulfotransferases:**

The crystal structures of SULTs, HSSTs, and TPSTs all share many central commonalities with respect to structure and function. All three families of sulfotransferases contain a conserved core structure consisting of 4-5 β-strands flanked on both faces with α-helixes (Figure 2A) (Gunal et al., 2019; Negishi et al., 2001; Teramoto et al., 2013). The need for binding specificity to PAPS, dictates structural conservation throughout the sulfotransferases. From the structural analysis of mouse estrogen sulfotransferase (mSULT1E1), two conserved motifs were identified (Figure 2) (Kakuta et al., 1997). The first strand-loop-helix motif is analogous to the P-loop structures found in the uridylate kinase family that cradles the 5’-phosphate of the PAPS, forming an extensive hydrogen bonding network via backbone amide interactions with the loop (referred to as the termed PSB-loop) (Kakuta et al., 1998a). Central to the loop is a conserved basic residue, usually a lysine but sometimes an arginine (such as in the TPSTs) that hydrogen bonds to the 5’-phosphate (Figure 2B-C) (Teramoto et al., 2013; Valentim et al., 2013). The second region, termed 3’SB, consists of a strand from the central β-sheet and an α-helix that runs across the top of the PSB loop that can form interactions with both the donor and acceptor substrates (Figure 2) (Kakuta et al., 1998a). This region is involved in binding of...
the 3′-phosphate of the PAP via sidechain interactions with a conserved arginine from the β-strand and a serine residue from the 3′SB α-helix (Figure 2B-C). The SULTs contain an additional conserved region of sequence GXXGXXK that as mentioned, was proposed to interact with the PAPS, based on its similarity in sequence to P-loops (Marsolais and Varin, 1995). The structure of mSULT1E1 revealed that indeed this region interacted with the PAPS, but via the 3′-phosphate and perhaps plays other critical roles in function, as discussed later.

The relative position of the substrate acceptor with respect to the position of the donor suggested an in-line displacement mechanism for the sulfo group transfer (Figure 3A-C) (Kakuta et al., 1997). This was further supported by a co-crystal structure of mSULT1E1 with vanadate and estradiol, as well as mSULT1E1 with PAPS alone (Figure 3D) (Kakuta et al., 1998b; Pedersen et al., 2002). Based on the structural comparisons to uridylylase kinase, it was hypothesized that the reaction would proceed via an S_N2-like associative reaction mechanism whereby the acceptor nucleophile could be primed by a catalytic base, allowing for nucleophilic attack on the sulfur of PAPS (Kakuta et al., 1997; Müller-Dieckmann and Schulz, 1994). The reaction would proceed through a trigonal bi-pyramidal transition state, with the leaving oxygen on the PAP and the incoming nucleophile in the axial positions (Figure 3B). This transition state was proposed to be mimicked by the crystal structure of mSULT1E1 in the presence of PAP and vanadate (Figure 3D) (Kakuta et al., 1998b). In the mSULT1E1/PAPS structure, the position of the conserved lysine on the PSB loop is found in a different orientation than when PAP is bound (Figure 3D) (Pedersen et al., 2002). Here, the lysine is found hydrogen bonding with the conserved serine from the 3′SB loop rather than with the 5′-phosphate of the PAP. It was proposed that this interaction discourages PAPS hydrolysis in the absence of acceptor substrate (Pedersen et al., 2002). Once the acceptor substrate is bound, the lysine was proposed to
undergo a conformational change to hydrogen bond with the phospho-sulfo bridging oxygen, functioning as a catalytic acid to encourage dissociation of the PAP leaving group. For the SULTs, a conserved histidine was identified within hydrogen bonding distance of the acceptor hydroxyl position that could function as a catalytic base to deprotonate the acceptor group (Figure 2B, Figure 3A-C) (Kakuta et al., 1997). Mutations of the conserved His108, Lys48 and Ser138 in mSULT1E1 all greatly reduced or eliminated detectable activity, supporting the proposed mechanism (Kakuta et al., 1998b; Pedersen et al., 2002).

One year after solving the structure of mSULT1E1, the Negishi lab solved the structure of the sulfotransferase domain (NST-1) of the bifunctional enzyme N-deacetylase/N-sulfotransferase isoform 1 (NDST-1) (Kakuta et al., 1999). This structure confirmed that conservation of the PAPS binding core for the sulfotransferases exists outside of the SULTs including the lysine from the PSB-loop and the arginine and serine from the 3’PB motif but lacked the conserved proposed histidine base (Figure 2C) (Kakuta et al., 1999). The proposed catalytic histidine was later found to be conserved in Golgi HS 2-OST and the 6-OST-1, but not in NDST-1, the 3-OSTs, or the TPSTs (Figure 2B-C) (Bethea et al., 2008; Edavettal et al., 2004; Kakuta et al., 1999; Liu et al., 2014; Moon et al., 2004; Moon et al., 2012; Tanaka et al., 2017; Teramoto et al., 2013; Xu et al., 2008; Xu et al., 2017b). In place of the histidine, these enzymes have a glutamate emanating from non-conserved structural elements within hydrogen bonding distance of the acceptor amine and hydroxyl atoms, respectively (Figure 2C) (Edavettal et al., 2004; Kakuta et al., 1999; Moon et al., 2004; Moon et al., 2012; Tanaka et al., 2017; Teramoto et al., 2013; Xu et al., 2008). Mutations of these proposed alternative catalytic base residues also greatly reduced activity (Edavettal et al., 2004; Kakuta et al., 2003; Teramoto et al., 2013; Xu et al., 2017b). Noteworthy is the fact that although the base may differ between different members
of the phylogenetic family tree, the relative orientations of the acceptor to the leaving group PAP are highly conserved, supporting an in-line transfer mechanism.

Interestingly, analysis using kinetic isotope effects and linear free-energy relationships on the SULTs, suggests a mostly dissociative $S_N1$-like reaction mechanism (Chapman et al., 2003; Hoff et al., 2006). This conclusion is supported by recent crystallographic analysis of the mouse SULT2A8 (Teramoto et al., 2021). This enzyme sulfates the $7\alpha$ position of bile acids. In this structure, the catalytic histidine is replaced with a leucine. Mutations to non-conserved His48 and Glu237, within proximity to the acceptor $7\alpha$-OH, show a greater impact on $K_M$ rather than $V_{\text{max}}$, indicating greater importance in binding than catalysis. These results suggest in sulfotransferases that the proposed catalytic base is required less for catalysis and more for binding and likely proper positioning. It is plausible that the ratio of dissociative to associative behavior of the sulfotransferases may vary within the family.

The SULTs are believed to employ a sequential mechanism whereby binding of both substrates occurs prior to product release (Leyh, 1993). Although the binding was originally thought to be through an ordered mechanism, recent more thorough analysis of the reaction, taking into account the formation of a dead-end product complexes, suggests the reaction may proceed via a random sequential mechanism (Wang et al., 2014a; Wang et al., 2014b; Zhang et al., 1998). These inhibitory dead-end complexes are present in many crystal structures and contain both the PAP product and the acceptor substrate bound (Table 1). The catalytic mechanism for SULT2A1 has been eloquently described to contain eight enzyme forms and 22 rate constants (Wang et al., 2014a). Their highly conserved sequence and structural features suggests this mechanism is maintained for many SULTs. While the reaction may proceed regardless of which substrate binds first, positive or negative cooperativity may regulate
substrate binding, suggesting a quasi-ordered mechanism may be a better description for some SULTs, depending on the enzyme and substrates examined (Tibbs et al., 2015). One example of this has been reported for SULT2A1 (Cook et al., 2012). Although SULT2A1 in the presence or absence of PAP does not demonstrate cooperative binding with DHEA, it does display negative cooperativity for Raloxifene in the presence of PAP (Cook et al., 2012). This has been suggested to be due to conformational changes associated with nucleotide binding to the SULT that restricts access of larger substrates such as Raloxifene to the acceptor binding site after PAPS binds. Thus, cellular levels of PAPS may contribute to a dynamic specificity of the enzyme (Tibbs et al., 2015).

Though sharing the in-line transfer geometry with the SULTs, an ordered mechanism as opposed to a bi-bi reaction mechanism has also been suggested for 6-OST-3, based on lack of binding to a substrate N-sulfoheparosan in the absence of PAPS (Sterner et al., 2014). Interestingly, based on LC mass spectrometry analysis, TPSTs were originally suggested to proceed through a two-site ping-pong reaction mechanism, whereby the PAPS and substrate bind independently and the reaction involves a covalent SO$_3$-histidine intermediate of the sulfotransferase prior to transfer of the SO$_3$ group to the acceptor tyrosine substrate residue (Danan et al., 2010). However, structural comparisons of the TPST-2 ternary complex structure with bound PAP and substrate peptide to those of the SULTs and HSSTs, along with the lack of a histidine proximal to the active site, support an in-line transfer. Additionally, like the 6OST-3, PAPS/PAP was required for binding of protein to peptide column, suggesting an ordered sequential mechanism of substrate binding (Niehrs and Huttner, 1990).

**Substrate specificity of SULTs**
While the SULTs, TPSTs and HSSTs all show conservation in nucleotide binding and catalytic mechanism, due to the wide ranges in shape, size, and charge of substrates, they can vary greatly in acceptor substrate recognition. The acceptor substrates of sulfotransferases can range in size from small phenolic compounds to very large protein or proteoglycan substrates. Modes of recognition and substrate orientation inside the binding pocket can also vary within the families themselves. Due to the smaller, often hydrophobic nature of the substrates, the SULTs’ acceptor substrate binding pockets are typically buried within the core of the protein by three loops (Figure 4a) (Tibbs et al., 2015). Originally termed phenol sulfotransferases, the human SULT1 family possess a substrate gate consisting of aromatic residues (Phe80 and Phe141 in hSULT1E1) with the one exception being SULT1B1 which contains a methionine at the first position (as seen in Tibbs et al Figure 7 (Tibbs et al., 2015)). These residues lie above and below the aromatic plane of the substrate and select for a phenolic acceptor, yet the rest of the pocket can accommodate diverse structural features (Figure 4A) (Duffel, 2010 updated 2016; Petrotchenko et al., 1999). The SULT2 family were originally referred to as hydroxysteroid or bile acid sulfotransferases but were later reclassified with the more general term alcohol sulfotransferases due to the developing knowledge of their broad specificity (Duffel, 2010 updated 2016; Lyon and Jakoby, 1980). The residues at the substrate gate differ in the SULT2 family allowing for non-aromatic acceptors such as DHEA. A thorough analysis on specificity overlap among and within the SULTs has been previously reviewed in fine detail (Coughtrie, 2016; Dong et al., 2012; Duffel, 2010 updated 2016). In general, the acceptor specificity is largely determined by three loops with variable sequences that define and restrict the entrance to the substrate binding pocket. These loops are flexible and often disordered in SULT structures, particularly in the absence of substrate (Tibbs et al., 2015). In human SULTs, loop 1 is ~9 to 10
residues and highly variable in the SULT1 family and is 5-6 residues shorter in the SULT2 and SULT4 families (Dong et al., 2012; Tibbs et al., 2015). It has been postulated that a smaller loop 1 for the SULT2s allow the binding of larger substrates compared to SULT1 family members (Dong et al., 2012). Loop 2 is located after the SB-helix and can contribute to substrate interactions (Figure 4A). Loop 3, also referred to as “lid” or “cap”, is the largest of the loops and covers both the PAPS and acceptor substrate binding sites when substrates are bound and may regulate binding cooperativity for certain substrates such as Raloxifene to SULT2A1, as mentioned previously (Cook et al., 2012; Leyh et al., 2013). Ordering of loop 3 is due in part to PAPS binding, as the conserved GXXGXXK is located at the C-terminus of this loop and, along with the preceding conserved arginine, forms interactions with the 3’-phosphate of the PAPS (Figure 4A). In fact, there are relatively few crystal structures of sulfotransferases without nucleotide present. Binding of the nucleotide generally improves the thermostability and increases the likelihood of crystallization by decreasing surface heterogeneity via limiting loop 3 flexibility (Tibbs et al., 2015).

Single amino acid differences between isoforms SULT1A1 and SULT1A2 as well as SULT1A1 and SULT1A3 have been suggested to dictate specificity among these isoforms (Dajani et al., 1998; Dong et al., 2012; Lu et al., 2010). As well, loops 1-3 display structural plasticity within single isoforms to accommodate different acceptor substrates, accentuating the need for structures with multiple acceptors bound to better interrogate these enzymes for design of specific inhibitors (Dong et al., 2012).

Crystal structures of human SULTs revealed a conserved KXXXTVXXXE dimerization motif that presents an unusually small protomer interface (Figure 4B) (Petrotchenko et al., 2001; Weitzner et al., 2009). The N-terminal lysine of this motif overlaps with the C-terminal lysine of
the GXXGXXXK sequence involved in 3′-phosphate binding at the end of loop 3 (Figure 4). The dimerization motif is located distal from the acceptor binding site and appears important for enzyme stability, based on studies with SULT1A1 and 1B1, but may also regulate function (Lu et al., 2009; Tibbs and Falany, 2016). Half-site reactivity has been demonstrated for SULT1E1, SULT1A1 and SULT2A1, suggesting that substrate binding at one active site may influence binding to the other protomer (Sun and Leyh, 2010; Wang et al., 2014a; Wang et al., 2014b). The proximity of the dimerization domain to residues involved in PAP binding suggests a potential mode of regulation described in detail by Tibbs et al. (Tibbs et al., 2015). However, this theme may not be universal, as SULT1B1 forms homodimers but does not display half-site reactivity (Tibbs and Falany, 2016). Homo- and heterodimers of SULTs have been described in vivo (Heroux and Roth, 1988; Kiehlbauch et al., 1995). Heterodimers could provide a mechanism by which substrates of one SULT could affect sulfation of another. Also, SULT4A1 alone has been hypothesized to play important roles in neuronal development, by heterodimerizing with and regulating the activity of other SULTs (Idris et al., 2020).

Substrate specificity of TPSTs

The TPSTs have broad overlapping substrate specificities and are believed to sulfate up to 1% of all tyrosine residues in the eukaryotic proteome (Baeuerle and Huttner, 1985; Hille et al., 1990). The human TPSTs -1 and -2 bind to intrinsically disordered sequences containing a tyrosine flanked by acidic residues within five amino acids on both the N- and C- terminal sides (Hortin et al., 1986; Rosenquist and Nicholas, 1993; Teramoto et al., 2013). The crystal structures of the catalytic domains of TPST-1 and -2 reveal that the tyrosine acceptor is buried deep within a positively charged binding pocket, necessitating the need for the peptide to take on an intrinsically unfolded conformation (Figure 5a) (Tanaka et al., 2017; Teramoto et al., 2013).
The peptide adopts a sharply bent configuration with the bend's apex immediately N-terminal (-1 position) to the tyrosine acceptor (0 position). For TPST-1 binding to a substrate peptide, the tyrosine and aspartate at the -1 position appear to be the most critical residues for binding (Figure 5A) (Tanaka et al., 2017). Like the human SULTs, the TPSTs form functional dimers, however, the dimerization interface is very different. For TPSTs the dimerization interface buries ~25% of the surface area and is located at the acceptor substrate binding pocket (Figure 5) (Tanaka et al., 2017; Teramoto et al., 2013). Although not critical for catalysis, dimerization appears to have a role in substrate binding (Teramoto et al., 2013). Structurally located at the position of loop 1 in the SULTs are three helices that form the majority of the dimer interface (Figure 5B). Heterodimers between TPST1 and 2 have been reported in vivo, likely due to the high sequence conservation of this region (Hartmann-Fatu et al., 2015). This region also contributes to acceptor substrate binding interactions at each active site. Two conserved arginines (Arg102 and Arg106) from the first α-helix, shown to be important for binding in TPST2, interact with the substrate in the same protomer binding site, while Arg123 (not critical for catalysis from the other protomer) also simultaneously interacts with the substrate (Figure 5A) (Teramoto et al., 2013). SULT-equivalent loop 2 from TPST forms β-sheet like interactions with the substrate peptide when present and is disordered in the absence of acceptor substrate (Teramoto et al., 2013). Although the extent of the conformational flexibility of “loop 3” in TPST remains unknown, it appears to be involved in dimer formation, exclusion of water from the active site and PAPS binding via a serine residue that interacts with the 5’-phosphate (Figure 5A) (Teramoto et al., 2013). It has been suggested that the presence of PAPS enhances peptide binding by organizing the substrate binding site (Niehrs and Huttner, 1990). The ability to bind and sulfonate a variety of peptide sequences may also make the enzyme susceptible to
nonproductive binding, as demonstrated in the crystal structure of TPST-1, where the Gastrin peptide is interpreted to be present in both a productive and a non-productive binding orientation (Tanaka et al., 2017).

**Substrate specificity of HSSTs**

HSSTs are highly specific for which functional group becomes sulfated and on the types of modifications (sulfation or epimerization) on the HS that can be accommodated within the active site. Unlike the SULTs and TPSTs, these enzymes utilize an open binding cleft to bind the large, often anionic polysaccharide acceptor substrates. Historically, the study of these enzymes was hindered by the inability to obtain homogeneous substrates. However, with the advent of chemoenzymatic synthesis, which utilizes glycosyltransferase and sulfotransferase enzymes to produce homogenous HS with specific lengths, sequences, sulfation and epimerization modifications, co-crystal structures with acceptor substrates bound with chicken 2-OST, zebra-fish 6-OST-3, and 3-OST-1 and -3 (mouse and human, respectively) have been obtained (Liu et al., 2014; Moon et al., 2004; Moon et al., 2012; Wander et al., 2021; Xu et al., 2017b). This structural information has provided a wealth of information into features dictating the specificity between families and isoforms within families.

Despite lacking an acceptor substrate, the crystal structure of the sulfotransferase domain of NDST-1 (NST-1) revealed a large open cleft that extends across the PAPS binding site to accommodate the acceptor substrate binding (Figure 6A) (Kakuta et al., 1999). The NDSTs’ sulfotransferase domains are most similar to that of the 3-OSTs (discussed later), with NST-1 sharing 28% sequence identity in the catalytic domain (Edavettal et al., 2004; Moon et al., 2004; Xu et al., 2008). Comparing NST-1 to the crystal structure of 3OST-1 revealed the acceptor binding cleft to be less positively charged than that of the 3OST-1 consistent with its substrate
containing a reduced anionic charge due to its earlier role in the HS maturation process (Edavettal et al., 2004). To determine residues involved in substrate binding, an acceptor substrate was modeled into the active site (Kakuta et al., 2003). Mutations were then made to residues lining the cleft on the 3’SB helix as well as a loop coined “sweet hill” with a unique insert of residues (640-649) on the opposite side of the cleft. Residues from this motif, including the proposed catalytic base Glu642, were shown to be important for activity (Kakuta et al., 2003).

The product of the N-deacetylase/N-sulfotransferase and C$_5$ epimerization becomes the substrate for 2-OST. To obtain the crystal structure of chicken 2-OST (92% sequence identity to human), 2-OST was crystallized as an MBP-fusion to enhance solubility and likelihood of crystallization (Bethea et al., 2008). Unlike the other HSSTs that appear to function as monomers, the structure revealed that 2-OST exists as a homotrimer with ~24% of each protomer’s surface area buried at the interface with the C-terminus of one molecule contributing to the substrate binding pocket of the other (Figure 6B-C). Removal of this tail disrupts trimer formation and greatly diminishes activity, supporting an important role for trimerization (Bethea et al., 2008). The active sites of each protomer are located ~ 50 Å apart and are believed to function independently of one another (Liu et al., 2014). The structure revealed extensive interactions with five saccharides in the bound heptasaccharide substrate. The N-sulfo groups on the GlcNSs flanking the acceptor uronic acid form multiple interactions, suggesting a GlcNS-containing pentasaccharide represents the minimum binding motif required for activity (Liu et al., 2014). The structure also revealed that sulfation on the 6-OH, of the adjacent GlcNS on the reducing end would result in significant steric clashes, while 6S on the nonreducing side could create electrostatic repulsion with the Glu349 on the neighboring protomer (Figure 6C).
together, the structures support the central dogma that 2-O-sulfation occurs after N-sulfation and before 6-O-sulfation. While 2-OST can transfer the sulfo group to both GlcA and IdoA saccharides, in a heptascaccharide with a mixture of GlcA or IdoA flanked by GlcNS, IdoA2S was produced 10:1 over GlcA2S (Liu et al., 2014). IdoA is typically found in $^1C_4$ or $^2S_0$, while GlcA adopts the $^4C_1$ conformation, suggesting that each would need to bind differently to the active site. However, the crystal structure, combined with mutagenesis, revealed that Arg189 may act to stabilize IdoA in the unexpected $^4C_1$ conformation, allowing both IdoA and GlcA to utilize the same active site in similar conformations (Figure 6C) (Liu et al., 2014). In contrast, a recent molecular modeling study suggest that for longer endogenous substrates, the IdoA would be in the preferred $^1C_4$ conformation and that Arg189 sterically excludes GlcA generating the preference for IdoA substrate (Gesteira et al., 2021).

Like 2-OST, the zebrafish 6-OST isoform 3 (zf6-OST-3; >70% sequence identity to human isoforms) was crystallized as an MBP-fusion (Xu et al., 2017b). Three crystal structures were obtained with three oligosaccharides (a hepta- and two hexasaccarides) comprised of different sequences that bound in similar conformations. While the 6-OSTs share many features with the other HSSTs, the orientation of substrate binding is completely different, with the substrate oriented almost perpendicular to that observed in the 2-OST and 3-OST crystal structures (Figure 6D). In all three 6-OST structures, the non-reducing end GlcNS is positioned with the 6-hydroxyl in the catalytic position. The structures revealed that the reducing end of the 2-OST and 3-OST cleft is occluded by the loop immediately following the 3′SB helix, structurally equivalent to that of loop 2 in the SULTs (Figure 6D). The 6-OSTs are the most promiscuous of the HSSTs with 6-OST-1 requiring only a trisaccharide composed of two glucosamines (GlcNS or GlcNAc) and a central uronic acid (GlcA, IdoA or IdoA2S) (Liu and
Liu, 2011). The lack of specificity is explained by the crystal structure, which shows that the majority of the interactions are with the acceptor GlcNS and the -GlcA-GlcNS- on its reducing side (Xu et al., 2017b). The NS on the acceptor GlcNS forms multiple interactions with the occluding “loop 2,” consistent with the preference of GlcNS substrate over GlcNAc. The lack of strict specificity by these enzymes may be important to generate a larger pool of diverse HS. 6OSTs are capable of sulfating on adjacent GlcNS-GlcA/IdoA repeat sequences. While all the substrates in the crystal structures bind across the reducing end of the binding interface, one of the substrates has an additional GlcA on the non-reducing side of the acceptor GlcNS suggesting the path of the non-reducing end, if it were present, would extend across the PAPS binding site (Figure 6D).

Humans have three isoforms of 6OST and the extent of specificity differences among them has been up for debate (Habuchi et al., 2000). Mapping the conserved residues of the three human isoforms onto the zf6-OST-3 revealed that all residues interacting with the substrate were conserved. This, combined with recent results of activity between the isoforms on different homogeneous substrates, supports similar specificities between the isoforms (Xu et al., 2017b). The 3-OSTs appear to have the strictest specificity requirements of all the HSSTs (Liu and Pedersen, 2007; Shworak et al., 1999). Of the seven human isoforms, the specificity of three (3-OST-1, 3-OST-3 and 3-OST-5) have been well characterized and examined structurally (Moon et al., 2004; Moon et al., 2012; Wander et al., 2021; Xu et al., 2008). 3-OST-1 is responsible for producing the pentasaccharide sequence GlcNS6S-GlcA-GlcNS6S3S-IdoA2S-GlcNS6S with strong anticoagulant activity (Liu and Pedersen, 2007; Liu et al., 1996). Alternatively, 3-OST-3 is involved in the production of the HS co-entry receptor for HSV-1, with the petasaccharide sequence GlcNS6S-IdoA2S-GlcNS6S3S-IdoA2S-GlcNS6S (Shukla et al., 1999). 3-OST-5 is
more promiscuous and can generate both substrates (Xia et al., 2002). Crystal structures exist for all three of these enzymes in binary complexes with PAP, and acceptor substrate-bound structures exist for 3-OST-1 and 3-OST-3 (Edavettal et al., 2004; Moon et al., 2004; Moon et al., 2012; Wander et al., 2021; Xu et al., 2008). Major differences in substrate specificities exist between the isoforms. 3-OST-1 prefers substrates that are 6-O-sulfated and contain a GlcA on the non-reducing side of the acceptor GlcNS, while 3-OST-3 prefers substrates lacking 6S that contain Ido2S saccharides flanking the acceptor GlcNS on both sides (Wang et al., 2017). Although substrates for both enzymes occupy the similar substrate cleft as found in NST and 2-OST, the conformations of the uronic acids differ in 3-OST-1 and -3 resulting in different interactions with the enzyme (Figure 7A). For 3-OST-1 binding, the GlcA and IdoA2S are in the $^4C_1$ and $^1C_4$ conformation respectively (Moon et al., 2012). For 3-OST-3, both IdoA2Ss are found in the $^2S_0$ conformation (Figure 7A) (Wander et al., 2021). Lys259 in 3-OST-3, not conserved in 3-OST-1, along with a Na$^+$ ion binding site on the reducing side have been suggested to contribute to 3-OST-3’s specificity (Figure 7A) (Moon et al., 2004; Wander et al., 2021). In addition, Lys259 may help to alleviate charge repulsions between the sulfo group of the Ido2S and the carboxylate group of the Ido2S on the non-reducing and reducing sides of the acceptor GlcNS, respectively. 3-OST-1 lacks these features contributing to its inability to accommodate a IdoA2S on the non-reducing side of the acceptor.

Based on the structures, a substrate ‘gate’ on the nonreducing side of the substrate binding cleft has been identified (Figure 7A) (Xu et al., 2008). Having a $^4C_1$ GlcA (3-OST-1) versus a $^2S_0$ IdoA2S (3-OST-3) on the non-reducing side of the acceptor GlcNS results in a different trajectory of the oligosaccharide through the gate for each substrate, suggesting that the
different sequences for the isoforms at the gate also contributes to specificity, which has been supported with mutagenesis studies (Wander et al., 2021; Xu et al., 2008).

While the 6S on two of the three ordered GlcNS6S saccharides in the 3-OST-1 heptaasaccharide substrate forms interactions with the enzyme, the 6S on the acceptor glucosamine is solvent-exposed (Moon et al., 2012). Due to the decrease in activity for 3-OST-3 on 6-O-sulfated substrates, it was originally hypothesized that 6S would generate steric and/or electrostatic clashes while binding to 3-OST-3. However, a crystal structure at 1.55 Å with two 3-OST-3 molecules in the asymmetric unit demonstrated that a 6S-containing octasaccharide was easily accommodated and, surprisingly, could bind in the same conformation as the octasaccharide lacking 6S (Figure 7B) (Wander et al., 2021). In contrast, the other 3-OST-3 molecule in the asymmetric unit bond the 6S-containing octasaccharide in a nonproductive manner, binding with the opposite polarity across the binding cleft (Figure 7B, yellow molecule). The structural information led to further analysis revealing that the 6S-containing substrate bound with 10-fold higher affinity and displayed a mix of product and substrate inhibition contributing to the perceived specificity difference (Wander et al., 2021).

Understanding disease through sulfotransferase structures

Missense genetic polymorphisms in sulfotransferases have been linked to differential metabolism of hormones and drugs by the SULTs as well as effecting activity of the HSSTs (Kurogi et al., 2021; Reuter et al., 2014; Schneeberger et al., 2020; Tornberg et al., 2011). These polymorphisms have been linked to diseases and conditions including malaria, cancer, congenital ichthyosis, idiopathic hypogonadotropic hypogonadism and other developmental disorders including neurological, skeletal and renal abnormalities (Table 2). Crystal structure analysis allow for hypothesizing of how these changes may affect function. While some of the
polymorphisms are located distally from the active site and may reflect changes in the protein
stability or interactions with binding partners, many of these mutations have been linked to the
substrate binding loops 2 and 3 for SULTs 1A1, 1B1, 1E1 and 2B1 and could impact substrate
binding and/or catalysis (Chung et al., 2009; Cohen et al., 2009; Heinz et al., 2017; Kinnersley et
al., 2016). Variants associated with disease also have been found lining the substrate binding
clefs of NDST-1, 2-OST-1, and 6-OST-1 (Najmabadi et al., 2011; Reuter et al., 2014;
Schneeberger et al., 2020; Tornberg et al., 2011). Disease associated variants are also found
clustered within the highly conserved 3′SB and 5′PSB strand-loop-helix motifs associated with
PAPS and acceptor substrate binding (Reuter et al., 2014; Schneeberger et al., 2020; Tibbs et al.,
2018). One of the variants for NDST-1, involves the proposed catalytic base, while the R189S
of 2-OST involves the residue proposed to dictate specificity for IdoA over GluA (Reuter et al.,
2014; Schneeberger et al., 2020). The R189S mutation would likely not inhibit GlcA binding but
would lose reactivity to IdoA as suggested by the R189A in vitro studies (Liu et al., 2014).

**ST structures in molecular modeling**

The value of sulfotransferase crystal structures has extended beyond basic understanding
of the enzymatic functions, serving as a starting point for in silico calculations to predict
substrates and inhibitors and better understand how conformational dynamics affect substrate
binding. Early work applied crystal structures and homology models to an adaptation of the
three-dimensional quantitative activity relationship (QSAR), referred to as comparative
molecular field analysis (CoMFA), to better understand enzyme kinetics and help predict
specificity of rat ASTIV and SULT1A3 (Sharma and Duffel, 2002; 2005; Sipilä et al., 2003). In
silico docking based on crystal structures was used to determine that PAPS binding can regulate
binding of larger substrates such as Raloxifene via conformational dynamics of loop 3 (Cook et
QSAR was later combined with molecular dynamics (MD) to consider both ligand and protein flexibility to successfully identify ligands of isoforms SULT1A1, SULT1A3 and SULT1E1 (Martiny et al., 2013). Others have utilized MD to study thermal stability and conformational flexibility associated with ligand binding in SULT2A1 (Zhao et al., 2016; Zhao et al., 2015; Zhou et al., 2019; Zhu et al., 2019). Utilizing molecular dynamics and an FDA-approved small molecule drug database, Cook et al were able to accurately predict known substrates and identify novel substrates for SULT1A1 and SULT2A1 (Cook et al., 2013b). Further MD analysis of SULT1A1 revealed a molecular clamp by two aromatic residues that reposition to “sandwich”, in a catalytically relevant position, the phenolic substrates that display enhanced affinities for binding in the presence of saturating nucleotide (Cook et al., 2015b). In contrast, substrates unaffected by nucleotide binding tend to “wander” in the binding pocket, rarely sampling a catalytically relevant position (Cook et al., 2015b). Analysis using MD simulations was also used to understand substrate inhibition differences between mouse and human SULT1E1, suggesting that a sub-pocket within the binding pocket of hSULT1E1 could accommodate the substrate in a non-catalytic binding mode (Stjernschantz et al., 2010). A recent approach in MD applied to SULT1A1, that includes the use of excited normal modes (MDeNM) has been used to sample more conformational space than classical MD and may provide more insight into substrate and/or inhibitor binding (Dudas et al., 2021).

Molecular docking experiments have also been used to better understand how allosteric interactions can regulate sulfotransferase function. SULT1A1 and 1A3 have been shown to contain two allosteric binding sites (Figure 8) (Cook et al., 2015a; Cook et al., 2017). Utilizing the crystal structure of SULT1A1 as a starting model of SULT1A3, monoamine neurotransmitter metabolites were screened against the known catechin-binding site of SULT1A3 predicting...
tetrahydrobiopterin (THB), an essential co-factor in monoamine neurotransmitter biosynthesis, as a potential allosteric regulator of SULT1A3 activity. When tested, THB had a $K_i=23\text{nM}$ for SULT1A3, but with no detectable affinity for SULT1A1, SULT1E1, SULT2A1 or SULT2B1 (Cook et al., 2017). Binding to the catechin-binding site, located at the entrance to the substrate binding pocket between loops 1 and 3, was confirmed through mutagenesis and spin-labeled triangulation NMR experiments (Cook et al., 2017). In 2019, Darrah et al identified a SULT1A3 specific inhibitor which also inhibited the sulfotransferase activity at nanomolar concentrations ($K_i=\sim34\text{nM}$). Using the spin-labeled triangulation NMR, combined with distance constraint molecular dynamics of models derived from crystal structures, a novel allosteric site specific to SULT1A3 was identified and confirmed by mutagenesis of surrounding residues (Darrah et al., 2019). This site is located outside the active site prior to loop 3, in a region disordered in the crystal structure of SULT1A3 when no substrates are present, that must open and close during catalysis to bind and release substrate (Figure 8) (Bidwell et al., 1999; Wang et al., 2014a). A subsequent study identified two other structurally similar compounds that also inhibited with nanomolar $K_i$s and were specific to SULT1A3 when compared to SULT1A1, SULT1E1 and SULT2A1. These compounds also inhibited SULT1A3 in cultured cells (Cook et al., 2021). Further molecular dynamics analysis supported the assessment that these compounds inhibit via loop 3 stabilization (Cook et al., 2021).

Modeling and molecular dynamics studies of HS substrates using the crystal structure of NST-1 with bound PAP has provided insight into residues that are involved in substrate recognition and influence catalysis (Gesteira et al., 2013; Gorokhov et al., 2000; Kakuta et al., 2003). Molecular dynamics was also utilized to provide information on global dynamics, residues involved in binding and a possible route for HS during modification by NDST-1 and 2-
OST (Gesteira and Coulson-Thomas, 2018). Conformational dynamics has also been investigated for TPST-2 (Singh et al., 2016a; Singh et al., 2016b).

In addition to these dynamic studies, quantum mechanics calculations as well as Quantum Mechanics/Molecular Mechanics (QM/MM) studies have been performed on SULT1E1, 3-OST-3, 2-OST and 6-OST to better understand the underlying catalytic mechanisms (Bartolotti et al., 1999; Gesteira et al., 2021; Lin et al., 2006; Sousa et al., 2016). QM/MM has also been applied to investigate the potential endocrine disrupting effects of hydroxylated bromodiphenyl ethers from environmental exposures via inhibition of SULT1A1 (Ma et al., 2021).

**Future Directions: Utilizing sulfotransferase structures for the development of new pharmacological tools**

Crystal structures can be powerful aids in the development of novel therapeutics. The Leyh group has utilized structural information on the closing of loop 3 in the presence of PAPS to design compounds that can interact with receptors but cannot be effectively sulfated by SULTs, which should increase the half-life of compounds *in vivo* (Cook et al., 2016). Such work could be useful for improved birth control bioavailability or treatment of Parkinson’s disease (Cook et al., 2016). Another example is the prodrug Oxamniquine, used to treat schistosomiasis which becomes activated through sulfation by cytosolic sulfotransferases. Resistance has emerged due to sulfotransferase variants. Crystal structures of these helminth sulfotransferases are being evaluated to help design new broad-spectrum oxamniquine derivatives that kill both *S. mansoni* and *S haematobium* species (Guzman et al., 2020; Valentim et al., 2013).
Historically, the design of effective therapeutic inhibitors has been hampered by obtaining a high degree of specificity for individual SULTs. The recent discovery of a selective allosteric site in SULT1A3 has changed that (Cook et al., 2021). SULT1A3 is the isoform responsible for sulfating 80% of serotonin levels in the brain resulting in decreased levels of the active neurotransmitter. Regulating unmodified serotonin levels is important for the treatment of depressive disorders. Molecular dynamics of SULT1A3 structural models, in combination with spin-labeled triangulation NMR, appears to have been critical for determining the allosteric binding pocket specific for SULT1A3. Taking advantage of ways to specifically inhibit SULT1A3 should greatly aid current and future therapeutics studies (Cook et al., 2021).

High-throughput screening has recently been used to search for inhibitors of the Golgi-resident TPST and GAG sulfotransferases (Byrne et al., 2018a; Byrne et al., 2018b; Cheung et al., 2017). The positive hits were then evaluated by docking to crystal structures to validate their ability to bind to the active site (Byrne et al., 2018a; Byrne et al., 2018b). Crystal structures of the TPSTs and HS 2-OST are also being utilized for structure-based design of nucleotide analog inhibitors (Kershaw et al., 2019). Such initial inhibitors could be used to improve specificity to TPSTs or 2-OST to probe function or potentially modulate disease states such as viral infection and inflammation (Byrne et al., 2018a; Byrne et al., 2018b; Kershaw et al., 2019).

A trending method for generation of pharmacological tools to probe physiological and pathophysiological pathways and potential therapeutics is utilization of a chemoenzymatic approach to synthesis homogenous GAG oligosaccharides. This method makes use of recombinant glycosyltransferases and sulfotransferases to efficiently produce the desired product, in contrast to tediously purifying heterogeneous compounds from source (Xu et al., 2017a; Xu et al., 2011; Zhang et al., 2020). Currently this process is being utilized for the
development of improved anticoagulant and autoinflammatory therapeutics to treat acute liver failure due to acetaminophen overdose (Arnold et al., 2020; Wang et al., 2021; Xu et al., 2014; Xu et al., 2017a).

Based on crystal structures of sulfotransferases engaged with oligosaccharide substrates, some of these enzymes have been engineered via site-directed mutagenesis to alter their specificities to improve control of the chemoenzymatic pathway (Liu et al., 2014; Liu et al., 2012; Xu et al., 2008; Xu et al., 2017b; Yi et al., 2020). Particularly, a triple mutation of zf6-OST-3 designed based on the crystal structures, allows for sulfation only on the non-reducing terminal glucosamine, providing fine-tuned control of 6-O-sulfation in the chemoenzymatic synthesis pathway (Yi et al., 2020). Cytosolic sulfotransferases are also being investigated for their ability to generate sulfated products for biological reagents and drug discovery (Lamb et al., 2006; Matsumura et al., 2018; Shi et al., 2012; Shimohira et al., 2017; Xie et al., 2020).

Historically, two obstacles to using sulfotransferases for product development were product inhibition by PAP and the cost of PAPS. Recycling PAP efficiently circumvents both obstacles by utilizing the cytosolic sulfotransferase rat ASTIV to regenerate PAPS from PAP and p-nitrophenyl sulfate (Burkart et al., 2000; Burkart et al., 1999). In concert with production of the desired sulfated product, recycling of PAP reduces both the need for additional PAPS and reduces PAP inhibition by keeping the PAP concentration low (Burkart et al., 2000; Burkart et al., 1999; Chen et al., 2005; Datta et al., 2020). Recent work has sought to improve upon the efficiency of this approach by re-engineering ASTIV (Zhou et al., 2019). Information obtained from the crystal structure of SULT1A1 was utilized to choose the loop sequence between Lys81 and Thr95 for molecular evolution that identified the double mutant L89S/E90L to have improved efficiency. These residues were then further subjected to site-saturation mutagenesis to
systematically test for improved efficiency. Zhou et al identified the variant L89M/E90Q, which has a 2-fold lower $K_M$ and a 10-fold higher catalytic efficacy than the wildtype enzymes (Zhou et al., 2019). This mutant could prove useful to improve chemoenzymatic synthesis by sulfotransferases. Thus, combining engineered SULTs with engineered GAG sulfotransferases could be useful in the discovery and production of pharmacological tools for probing pathways and designing therapeutics.

**Conclusion**

Crystal structures of the SULTs, HSSTs and TPSTs have provided a crucial foundation of structural information to better understand the mechanisms and specificities of these enzymes. These structures are proving useful for *in-silico* substrate and inhibitor screening from chemical databases as well as for inhibitor design. Mutational engineering based on the sulfotransferase structures has proven successful in altering specificity and kinetic properties. Understanding how the sulfated products of these enzymes interact with their downstream receptors and targets will likely stimulate an interactive, iterative process of redesigning the specificity of the sulfotransferases to enhance specificity of the products with their targets. The design of inhibitors along with the use of sulfotransferases in synthesizing compounds should enhance the production of pharmacological tools for probing pathways and the design of novel therapeutics.

**Figure legends:**
Figure 1) Sulfotransferases Covered in this review. Sulfotransferases transfer a sulfo group ($\text{SO}_3^-$) from PAPS, which is generated in the cytosol by the bifunctional PAPS synthases, to many different types of acceptor substrates. The SULT enzymes in the cytosol sulfate endogenous and exogenous small molecules, while the Golgi-associated sulfotransferase sulfate macromolecules such as proteins and glycosaminoglycans.

Figure 2) Representative structures of the different types of sulfotransferases and their PAPS Binding Sites. A) From left to right: crystal structure of hSULT1E1 in complex with PAP and estradiol (PDB code 4JVL) (Gosavi et al., 2013); crystal structure of 3-OST-3 with PAP and an 8mer NS2S heparan sulfate bound (PDB code 6XL8) (Wander et al., 2021); crystal structure of TPST-1 with PAP and a polypeptide substrate bound (PDB code 5WRI) (Tanaka et al., 2017). The PAP is colored green and acceptor substrates are cyan. The strand-loop-helix containing the PSB loop is colored pink while the strand-loop-helix containing the 3′-phosphate binding motif (3′SB) is colored light purple. The remaining two strands making up the central β-sheet are colored light green. B) Comparisons of the PAP binding motifs and catalytic residues. The SULTs including SULT1E1 (green) and heparan sulfotransferases 2-OST(yellow) and 6-OST(pink) all utilize a lysine on the PSB loop and histidines for the proposed catalytic base (PDB codes 4JVL, 4NDZ, and 5T0A respectively) (Gosavi et al., 2013; Liu et al., 2014; Xu et al., 2017b). The acceptor hydroxyl on the substrates all superimpose well supporting a conserved mechanism. C) Comparisons of the PAP binding motifs and catalytic residues of the sulfotransferase domain of SULT1E1 (green), NDST-1 (gray), 3-OST-3 (wheat), and TPST-1 (light purple) (PDB codes: 4JVL, NST1, 6XL8, and 5WRI respectively) (Gosavi et al., 2013; Kakuta et al., 1999; Tanaka et al., 2017; Wander et al., 2021). NDST-1, 3OST-3 and TPST-1 appear to utilize a conserved glutamate for their catalytic base, as opposed to the SULTs, 2OST
and 6OST. However, the PAPs and acceptor hydroxyls on the substrates all superimpose perfectly, supporting a conserved in-line reaction mechanism for all the vertebrate sulfotransferases. For panels B and C, only the acceptor saccharide of the substrate is shown for 3-OST-3 and 2-OST and acceptor tyrosine of the peptide for TPST-1. No substrate is present in the crystal structure of NST-1.

**Figure 3) Catalytic mechanism of the sulfotransferases.**

A-C) Proposed catalytic mechanism of the PAPS-dependent sulfotransferases. D) Common architecture of the sulfotransferase PAPS binding site including the PSB-loop and 3′SB strand-loop-helix from the crystal structure of mSULT1E1 (PDB code 1BO6) with PAP(white) and a vanadate molecule (green) in a trigonal bi-pyramidal arrangement thought to mimic the transition state. Also shown are the proposed catalytic base (His) and acid (Lys) (Kakuta et al., 1998b). Superimposed on to this structure are PAPS (cyan) and estradiol (pink) from two different structures of human SULT1E1 (PDB codes 1HY3 and 4JVL respectively) with their equivalent His and Lys residues (Gosavi et al., 2013; Pedersen et al., 2002). When PAPS is present, the lysine forms a hydrogen bond with a conserved serine from the 3′SB binding region but is bound to the 5′phosphate when PAP is present. These structures, combined with comparisons to uridylate kinase, shaped the hypothesis for the reaction mechanism shown in panels A-C.

**Figure 4) Substrate Binding and Dimerization of SULTS.** A) A representative SULT crystal structure, SULT1E1 with PAP (green) and estradiol (cyan) bond (PDB code 4JVL) (Gosavi et al., 2013). Shown are Loops 1 (yellow), -2 (magenta) and -3 (green) that contribute to substrate
specificity and binding. The SULT-specific GXXGXXK region that connects loop 3 to the dimerization domain (orange) is colored lemon. The acceptor hydroxyl of the estradiol is designated with a red asterisk. Also shown are highly conserved aromatic residues (wheat) found in the SULT1 enzymes that contribute to selectivity for phenolic compounds.

Figure 5) Substrate Binding and Dimerization of TPST-1. A) Substrate binding site of TPST-1 (PDB code 5WRI) (Tanaka et al., 2017). Residues Arg102 and Arg106 from the α-helix bundle and “loop 2” (magenta) contribute significantly to substrate binding. Residues from “loop 3” (green) form interactions with both the acceptor (Arg285) and donor substrates (Ser286). Arg123 from the other protomer forms a non-essential interaction with the substrate. The acceptor tyrosine is positioned for catalysis and forms a hydrogen bond with the proposed catalytic base E100 (red dashed line). The acceptor hydroxyl on the tyrosine and the Asp at the -1 position of the substrate are designated with red and blue asterisks, respectively. B) Dimer interface of TPST-1. One protomer is colored light purple while the other is colored grey with the equivalent to loops 1,2, and 3 of the SULTs colored in yellow, magenta, and dark green respectively.

Figure 6) Substrate Binding of the Heparan Sulfate Sulfotransferases. A) Crystal structure of the sulfotransferase domain of NDST-1 (PDB code 1NST) with PAP shown (Kakuta et al., 1999). The regions colored in light purple (α-helix from 3′SB motif) and wheat (“Sweet Hill” region, including Glu642, the catalytic base) have been shown to be important for substrate binding and lie along an open cleft across the active site, which is similar to that seen in 3OST-3.
(Figure 2A). B) Crystal structure of the 2OST trimer with protomers shown in white, light purple and pink (PDB code 4NZD). PAP and the octasaccharide substrate bound to each active site are shown in green and cyan, respectively. Of note, the C-terminal residues of one protomer extend into the active site of the other (Liu et al., 2014). C) Active site of 2OST suggests how the enzyme accommodates both IdoA and GlcA in the active site by supporting binding of a $^4C_1$ acceptor sugar conformation that relies on Arg189. The positions of the 6-OH as shown would likely exclude 6S moieties, due to steric conflict on the reducing end and possible electrostatic repulsion with Glu349 from another protomer on the GlcNS on the non-reducing side. The acceptor hydroxyl of the octasaccharide is designated with a red asterisk while the reducing end and nonreducing ends are labeled r and nr, respectively. D) Crystal structure of zf6OST-3 (PDB code 5T0A) active site (light purple) with PAP (green) and bound heptamer (wheat) substrate (Xu et al., 2017b). Hydrogen bonds with substrate are shown in black dashed lines. The acceptor hydroxyl is designated with a red asterisk. 2-OST (grey) is superimposed with its substrate (transparent cyan). The superposition reveals 6-OST binds its substrate with opposite polarity, relative to the active site, as compared to 2-OST and 3-OSTs. In 6-OST, a loop including T209 (magenta), blocks the cleft found in the other heparan sulfotransferases and forms interactions with the N-sulfo moiety on the acceptor glucosamine.

**Figure 7) Substrate Binding to 3-OST-1 and 3-OST-3 Active Sites.** A) Crystal structure of 3-OST-3 (pink) with bound NS2S 8mer oligosaccharide (cyan) (PDB code: 6XL8) superimposed with 3-OST-1(grey) with bound heptasaccharide (green) (PDB code: 3UAN) (Moon et al., 2012; Wander et al., 2021). Residue side chains that differ between the two isoforms lining the substrate binding pocket are drawn in stick. Hydrogen bonds are depicted with dashed black lines and interactions with the Na$^+$ ion (pink) involved in substrate binding to 3OST-3 are shown.
in solid-black lines. The positions of the two uronic acids flanking the acceptor glucosamine, the acceptor 3-OH on the glucosamine and the reducing (r) and non-reducing (nr) ends of the oligosaccharide are labeled. Residues associated with the non-reducing end “substrate gate” and labeled and denoted with blue arrows. B) Crystal structure of 3-OST-3 with the productive binding mode of NS2S6S containing 8mer bound to one protomer (olive) with the position of the NS2S6S containing 8mer in the nonproductive mode superimposed (all yellow) (PDB code 6XK6) (Wander et al., 2021). The nonproductive binding mode has reversed polarity with respect to the active site and the acceptor substrate is not in position for catalysis (magenta asterisk). The correct position for the acceptor 3OH hydroxyl is designated with a red asterisk. The NS2S substrate from the 3-OST-3 structure PDB code:6XL8 is also superimposed (cyan) and displays very similar binding to productive positioning of the NS2S6S.

**Figure 8** The Allosteric Binding Sites of SULT1A3. Model of SULT1A3 with bound PAPS (green) and inhibitor CMP-8 (coral) bound to an allosteric site via aromatic stacking with His226 and Phe222 (model found at: https://www.modelarchive.org/doi/10.5452/ma-qtj80) (Darrah et al., 2019). The position of the acceptor substrate dopamine (cyan) was superimposed from PDB coordinates 2A3R (Lu et al., 2005). Also shown are residues Tyr76 (orange), Asp86 and Glu89 (both yellow) that compose the catechin-binding allosteric site that binds the monoamine neurotransmitter cofactor tetrahydrobiopterin (Cook et al., 2017). Loops 1, 2 and 3 are colored yellow, magenta, and dark green respectively as seen in SULT1E1 (Figure 4). In the crystal structure of SULT1A3 without substrate present, residues 216-261 are disordered. This comprises loop 3 and the two helices (light green) N-terminal to loop 3 that make up one of the allosteric binding sites (PDB code 1CJM) (Bidwell et al., 1999).

**ACKNOWLEDGEMENTS**
We thank J. Krahn and L. Perera for critical reading of the manuscript and Dr. Masahiko Negishi for his mentoring and friendship as well as unwavering support of structural biology at the NIEHS/NIH.

AUTHOR CONTRIBUTIONS

Performed data analysis: L.C. Pedersen and M. Yi

Wrote or contributed to the writing of the manuscript: L.C. Pedersen, M. Y., L.G. Pedersen, A.M. Kaminski
Table 1. Representative Structures of the human SULT, TPST and HSSTs

<table>
<thead>
<tr>
<th>Sulfotransferase</th>
<th>Acceptor</th>
<th>Nucleotide</th>
<th>PDB code</th>
<th>Res Å</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human Cytosolic Sulfotransferases (SULTs)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SULT1A1</td>
<td>p-nitrophenol</td>
<td>PAP</td>
<td>1LS6</td>
<td>1.9</td>
<td>(Gamage et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>estradiol</td>
<td>PAP</td>
<td>2D06</td>
<td>2.3</td>
<td>(Gamage et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>p-nitrophenol</td>
<td>PAP</td>
<td>3QVU</td>
<td>2.5</td>
<td>(Alcolombri et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>3-cyanoumbelliferone</td>
<td>PAP</td>
<td>3QVV</td>
<td>2.35</td>
<td>(Alcolombri et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>2-naphthol</td>
<td>PAP</td>
<td>3U3K</td>
<td>2.36</td>
<td>(Berger et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>3-cyano-7-hydroxycoumarin</td>
<td>PAP</td>
<td>3U3M, 3U3O</td>
<td>2.3, 2.0</td>
<td>(Berger et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>p-nitrophenol</td>
<td>PAP</td>
<td>3U3R</td>
<td>2.36</td>
<td>(Berger et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>PAP</td>
<td>3U3J</td>
<td>2.70</td>
<td>(Berger et al., 2011)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>PAP</td>
<td>4GRA</td>
<td>2.56</td>
<td>(Cook et al., 2013a)</td>
</tr>
<tr>
<td>SULT1A1*3</td>
<td>-</td>
<td>PAP</td>
<td>1Z28</td>
<td>2.3</td>
<td>(Lu et al., 2003)</td>
</tr>
<tr>
<td>Gene</td>
<td>Substrate</td>
<td>PAP</td>
<td>Product</td>
<td>Assay</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
<td>-----</td>
<td>----------</td>
<td>---------</td>
<td>--------------------</td>
</tr>
<tr>
<td>SULT1A2</td>
<td></td>
<td>PAP</td>
<td>1Z29</td>
<td>2.4</td>
<td>(Lu et al., 2010)</td>
</tr>
<tr>
<td>SULT1A3</td>
<td>Dopamine</td>
<td>PAP</td>
<td>2A3R</td>
<td>2.6</td>
<td>(Lu et al., 2005)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>1CJM</td>
<td>2.4</td>
<td>(Bidwell et al., 1999)</td>
</tr>
<tr>
<td>SULT1B1</td>
<td>Resveratrol</td>
<td>PAP</td>
<td>3CKL</td>
<td>2.0</td>
<td>TBP</td>
</tr>
<tr>
<td>SULT1C2α</td>
<td></td>
<td>PAP</td>
<td>2Z5F</td>
<td>2.1</td>
<td>(Dombrovski et al., 2006)</td>
</tr>
<tr>
<td>SULT1C3d</td>
<td></td>
<td>PAP</td>
<td>2H8K</td>
<td>3.2</td>
<td>(Allali-Hassani et al., 2007)</td>
</tr>
<tr>
<td>SULT1C4β</td>
<td>Pentachlorophenol</td>
<td>PAP</td>
<td>2GWH</td>
<td>1.80</td>
<td>(Allali-Hassani et al., 2007)</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>2AD1</td>
<td>2.00</td>
<td>(Allali-Hassani et al., 2007)</td>
</tr>
<tr>
<td>SULT1E1</td>
<td>Estradiol</td>
<td>PAP</td>
<td>4JVL</td>
<td>1.94</td>
<td>(Gosavi et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>Tetrabromobisphenol A</td>
<td>PAP</td>
<td>4JVM</td>
<td>1.99</td>
<td>(Gosavi et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>3,5,3',5'-tetrachloro-biphenyl-4-4'-diol</td>
<td>PAP</td>
<td>1G3M</td>
<td>1.7</td>
<td>(Shevtsov et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>3-hydroxybromodiphenyl ether</td>
<td>PAP</td>
<td>4JVN</td>
<td>2.05</td>
<td>(Gosavi et al., 2013)</td>
</tr>
<tr>
<td>Enzyme</td>
<td>Substrate</td>
<td>Catalytic Site</td>
<td>K&lt;sub&gt;m&lt;/sub&gt; (mM)</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>SULT2A1</td>
<td>Androsterone</td>
<td>-</td>
<td>1.80</td>
<td>(Pedersen et al., 2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DHEA</td>
<td>-</td>
<td>2.70</td>
<td>(Chang et al., 2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3Beta,5alpha)-3-Hydroxyandrostan-17-one</td>
<td>-</td>
<td>1.99</td>
<td>(Rehse et al., 2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lithocholic acid</td>
<td>PAP</td>
<td>2.2</td>
<td>TBP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAPS</td>
<td>2.3</td>
<td>TBP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAP</td>
<td>2.4</td>
<td>(Pedersen et al., 2000)</td>
<td></td>
</tr>
<tr>
<td>SULT2B1a</td>
<td>-</td>
<td>PAP</td>
<td>2.91</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SULT2B1b</td>
<td>DHEA</td>
<td>PAP</td>
<td>2.5</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pregnenolone</td>
<td>PAP</td>
<td>2.3</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAP</td>
<td>2.4</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>SULT4A1</td>
<td>-</td>
<td>-</td>
<td>2.24</td>
<td>(Allali-Hassani et al., 2007)</td>
<td></td>
</tr>
</tbody>
</table>

**Tyrosyl Protein Sulfotransferase (TPSTs)**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cysteine Motif</th>
<th>Catalytic Site</th>
<th>K&lt;sub&gt;m&lt;/sub&gt; (mM)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPST-1</td>
<td>DFEDYEFDF</td>
<td>PAP</td>
<td>1.60</td>
<td>(Tanaka et al., 2017)</td>
</tr>
<tr>
<td></td>
<td>EEEEEEAYGWMDF</td>
<td>PAP</td>
<td>2.33</td>
<td>(Tanaka et al., 2017)</td>
</tr>
<tr>
<td>Heparan Sulfate Sulfotransferases (HSSTs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDST-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(sulfotransferase catalytic domain)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-OST&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-OST-3&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNS-GlcA-GlcNS-IdoA2S-GlcNS-GlcA-pNP</td>
<td>PAP</td>
<td>5T05</td>
<td>1.95 (Xu et al., 2017b)</td>
<td></td>
</tr>
<tr>
<td>GlcNS-GlcA-GlcNS-GlcA-GlcNS-GlcA-pNP</td>
<td>PAP</td>
<td>5T03</td>
<td>2.1 (Xu et al., 2017b)</td>
<td></td>
</tr>
<tr>
<td>GlcA-GlcNS-GlcA-GlcNS-GlcA-Glc-pNP</td>
<td>PAP</td>
<td>5T0A</td>
<td>1.95 (Xu et al., 2017b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-OST-1&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNAc6S-GlcA-GlcNS6S-IdoA2S-GlcNS6S-GlcA-UA</td>
<td>PAP</td>
<td>3UAN</td>
<td>1.84 (Moon et al., 2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-OST-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-OST-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNAc-GlcA-GlcNS-Ido2S-GlcNS-Ido2S-GlcNS-</td>
<td>PAP</td>
<td>6XL8</td>
<td>2.34 (Wander et al., 2021)</td>
<td></td>
</tr>
<tr>
<td>GlcA-pNP</td>
<td>PAP</td>
<td>6XKG</td>
<td>1.55</td>
<td>(Wander et al., 2021)</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>----------------------</td>
</tr>
<tr>
<td>δUA2S-IdoA2S-GlcNS6S-IdoA2S</td>
<td>PAP</td>
<td>1T8U</td>
<td>1.95</td>
<td>(Moon et al., 2004)</td>
</tr>
<tr>
<td>-</td>
<td>PAP</td>
<td>1T8T</td>
<td>1.85</td>
<td>(Moon et al., 2004)</td>
</tr>
<tr>
<td>3-OST-5</td>
<td>-</td>
<td>PAP</td>
<td>3BD9</td>
<td>2.3</td>
</tr>
</tbody>
</table>

a) In PDB under 1SULT1C1

b) In PDB under 1SULT1C2
c) *Gallus gallus*
d) *Danio rerio*
e) *Mus musculus*
f) TBP (To be published)
<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP cluster</th>
<th>Impact</th>
<th>Amino acid change</th>
<th>Location in Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SULT1A1</td>
<td>rs1042008</td>
<td>Increased risk of oral squamous cell carcinoma</td>
<td>His149Tyr</td>
<td>Loop 2</td>
</tr>
</tbody>
</table>

Table 2. Polymorphisms of human sulfotransferases with potential clinical significance
<table>
<thead>
<tr>
<th>Gene</th>
<th>Reference Number</th>
<th>Description</th>
<th>Associated Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SULT1A2</strong></td>
<td></td>
<td>rs1801030 Endometrial cancer (Rebbeck et al., 2006)</td>
<td>Met223Val&lt;sup&gt;a&lt;/sup&gt; Pre-loop 3</td>
</tr>
<tr>
<td></td>
<td>rs4987024</td>
<td>Inversely associated with bladder cancer risk (Figueroa et al., 2008)</td>
<td>Tyr62Phe Away from active site</td>
</tr>
<tr>
<td></td>
<td>rs11569729</td>
<td>Higher odds ratio in glioblastoma gliomas (Kinnersley et al., 2016)</td>
<td>Thr261Met Loop 3</td>
</tr>
<tr>
<td><strong>SULT1C4</strong></td>
<td>Rs1402467</td>
<td>Increased relapse rate in acute myeloblastic leukemia (Monzo et al., 2006)</td>
<td>Asp5Glu&lt;sup&gt;a&lt;/sup&gt; N-terminus not in structure</td>
</tr>
<tr>
<td><strong>SULT1E1</strong></td>
<td></td>
<td>Increased risk of breast cancer (Cohen et al., 2009)</td>
<td>His224Gln Pre-loop 3</td>
</tr>
<tr>
<td></td>
<td>rs140526640</td>
<td>May relate to autosomal-recessive congenital ichthyosis (Youssefian et al., 2019)</td>
<td>Glu78Lys&lt;sup&gt;a&lt;/sup&gt; Buried H-bond to Arg100</td>
</tr>
<tr>
<td></td>
<td>rs1303127476</td>
<td>May relate to autosomal-recessive congenital ichthyosis (Youssefian et al., 2019)</td>
<td>Arg100Trp&lt;sup&gt;a&lt;/sup&gt; Buried H-bond to Glu78</td>
</tr>
<tr>
<td><strong>SULT2B1</strong></td>
<td></td>
<td>rs114167424 May cause autosomal-recessive congenital ichthyosis (Heinz et al., 2017)</td>
<td>Pro149Arg 3'SB loop</td>
</tr>
<tr>
<td></td>
<td>rs1052131</td>
<td>Associated with esophageal squamous cell carcinoma risk (Hong et al., 2019)</td>
<td>Asp316Glu C-terminus not ordered</td>
</tr>
<tr>
<td></td>
<td>rs762765702</td>
<td>May cause autosomal-recessive ichthyosis (Heinz et al., 2017)</td>
<td>Arg274Gln Loop 3, binds 3' phosphate of PAPS</td>
</tr>
<tr>
<td><strong>NDST1</strong></td>
<td></td>
<td>rs606231456 Significant overlap in both demonstrated and apparent intellectual disability, muscular hypotonia, epilepsy, and postnatal growth (Najmabadi et al., 2011; Reuter et al., 2014)</td>
<td>Arg709Gln On 3'SB helix</td>
</tr>
<tr>
<td></td>
<td>rs606231457</td>
<td>Significant overlap in both demonstrated and Glu642Asp</td>
<td></td>
</tr>
</tbody>
</table>
apparent intellectual disability, muscular hypotonia, epilepsy, and postnatal growth deficiency (Reuter et al., 2014).

**Catalytic base on Sweet Hill**

- **rs606231458** hypotonia, epilepsy, and postnatal growth Phe640Leu deficiency (Reuter et al., 2014)

- **rs606231459** hypotonia, epilepsy, and postnatal growth Gly611Ser deficiency (Reuter et al., 2014)

**Associated with intellectual disability, muscular hypotonia, epilepsy, and postnatal growth deficiency (Reuter et al., 2014)**

**5’PSB-loop**

- **rs758990524** comprising neurological, skeletal, and renal Asp165Tyr abnormalities (Schneeberger et al., 2020)

- **rs1651972168** comprising neurological, skeletal, and renal Phe176Ser abnormalities (Schneeberger et al., 2020)

**3’SB loop**

- **HS2ST1**

- **rs60532842** Associated with P. falciparum parasitaemia Ala85Ser (Atkinson et al., 2012) N-terminal of sulfotransferase domain

- **rs9906590** Associated with P. falciparum parasitaemia Glu363Lys/Asp363 Glu363Gln Away from active site (Atkinson et al., 2012; Nguyen et al., 2018)

- **HS3ST3B1**

- **rs62056073** Associated with P. falciparum parasitaemia Ile196Val (Atkinson et al., 2012; Nguyen et al., 2018) Ile196Phe Away from active site

- **rs62636623** Associated with P. falciparum parasitaemia Gly83Arg (Atkinson et al., 2012; Nguyen et al., 2018) Gly83Trp N-terminal of sulfotransferase domain
- List was obtained by utilizing the NSBI dbSNP search with clinical significance (likely pathogenic, pathogenic, and pathogenic likely pathogenic) and function class (missense).

a) not found in NSBI dbSNP using listed criteria

b) HS2ST analysis based on crystal structure of chicken 2-OST PDB code: 4ndz

c) H3ST3B1 structural analysis based on structure of equivalent residue in H3ST3A1 PDB code: 6xl8

d) HS6ST1 and HS6ST2 structural analysis based on crystal structure of zebrafish 6-OST-3 PDB id code: 5t0A.

Numbering of Arg306, Arg323, Arg382 and Met404 are listed as Arg296, Arg313, Arg372 and Met294 in Tornberg et al (Tornberg et al., 2011).

| rs780352591 | Found in hypogonadotropic hypogonadism patients (Tornberg et al., 2011) | Arg306Trp<sup>1</sup> | Away from active site |
| rs201307896 | Found in hypogonadotropic hypogonadism patients (Tornberg et al., 2011) | Arg306Gln<sup>1</sup> | Away from active site |
| rs761325768 | Found in hypogonadotropic hypogonadism patients (Tornberg et al., 2011) | Arg323Gln<sup>1</sup> | Lines non-reducing end of substrate binding pocket |
| rs199538589 | Found in hypogonadotropic hypogonadism patients (Cangiano et al., 2019; Choi et al., 2015; Tornberg et al., 2011) | Arg382Trp | Extended C-terminal tail away from active site |
| rs866919041 | Associated with X-linked intellectual disability and severe myopia in two male twins (Paganini et al., 2019). | Gly306Cys | Disordered loop away from active site |

HS6ST1<sup>d</sup>

HS6ST2<sup>d</sup>
References:


Figure 1

**Sulfotransferases**

\[ \text{PAPS} + R-X \rightarrow \text{PAP} + RX - SO_3^- \]

\[ X = \text{-OH or -NH}_2 \]

**Golgi**

Tyrosyl Protein and HS Sulfotransferases (TPST-1 and -2; HSSTs)

Substrates:
proteins + glycosylaminoglycans

**Cytosol**

Cytosolic Sulfotransferases (SULTs)

Substrates:
- **Endogenous**: steroid and thyroid hormones, catecholamines, bile acids
- **Exogenous**: xenobiotics, environmental exposures
Figure 2

A  SULT1E1  3-OST-3  TPST-1

B  Arg  Ser  Lys  His

C  Lys/Arg  Glu
Figure 3
Figure 4
Figure 6
Figure 7
Figure 8