STRUCTURAL AND FUNCTIONAL ANALYSIS OF HEPARAN SULFATE SULFOTRANSFERASES

Heather Nicole Bethea

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Pharmaceutical Sciences (Medicinal Chemistry and Natural Products)

Chapel Hill
2010

Approved by:

Michael Jarstfer, Ph.D.

Jian Liu, Ph.D.

Lars Pedersen, Ph.D.

Scott Singleton, Ph.D.

Qisheng Zhang, Ph.D.

ABSTRACT

HEATHER BETHEA: Structural and Functional Analysis of Heparan Sulfate Sulfotransferases
(Under the direction of Jian Liu, Ph.D.)

Heparan sulfate (HS), a major polysaccharide component of the vascular system, is involved in regulating a number of functions of the blood vessel wall including blood coagulation, cell differentiation, and the inflammatory response. The wide range of biological functions makes HS an attractive therapeutic target. The long term goal of our research involves utilizing an enzyme-based approach to develop HS-based therapeutics for treating thrombotic diseases, cancer and excessive inflammatory responses. The biosynthesis of HS involves multiple specialized sulfotransferases such as 2-*O*-sulfotransferase (2OST) and 6-*O*-sulfotransferase (6OST), which are essential for preparing HS with activities in regulating vascular development and blood coagulation. The substrate specificity of the HS sulfotransferases controls the sulfation patterns of HS, permitting HS to exhibit a specific function, however, limited knowledge regarding the mechanism of these enzymes has hindered our ability to prepare functionally-specific HS. We aim to understand the mechanism of action of these two enzymes in hopes of developing heparin/HS with improved anticoagulant efficacy.

In this dissertation, we present successful crystallization of 2OST in complex with 3'-phosphoadenosine 5'-phosphate (PAP). The substrate recognition mechanism of 2OST was

examined by way of extensive structurally guided mutational analysis. Several residues were identified, including Arg-189, Tyr-94, and His-106, that are responsible for dictating the substrate specificity of 2OST. Despite success with the crystallization of 2OST, the crystallization of 6OST has been an ongoing process. A promising expression construct for crystallization purposes has been identified for 6OST-1. Using a homology model of 6OST-3 with structurally known 3OST-3, several residues involved in PAPS and substrate binding were proposed.

To my mother, Faye S. Barfield, whose everlasting love and support helped guide me throughout my educational endeavors.

To my fiancé, Joshua R. Horne, whose continuous love, support, and optimism made completion of this dissertation possible.

ACKNOWLEDGEMENTS

First and foremost, I would like to offer my sincerest gratitude to my advisor, Dr. Jian Liu, for his guidance throughout my graduate career. His encouragement, optimism, and support helped offer me the strength and perseverance to strive through completion of these studies over the past five years. Dr. Liu is an upstanding mentor who takes pride in the success of his students and I will always be grateful to him for helping me become a better scientist.

Next, I would like to acknowledge Dr. Lars Pedersen for assistance with the structural analysis of 2OST. I would also like to thank Dr. Pedersen for serving as a member of my dissertation committee, providing me with continuous counseling throughout completion of the crystallization studies for 2OST, 6OST, and TPST. I must thank Andrea Moon in Dr. Lars Pedersen's laboratory for teaching me how to manipulate the structural analysis software, PyMol, and Dr. Raj Gosavi for providing me with a library of TPST enzymes. To Dr. Jerry Turnbull, I would like to thank him for providing me with an amazing opportunity to study abroad at the University of Liverpool. I send my warmest appreciation to my other dissertation committee members, Dr. Mike Jarstfer, Dr. Scott Singleton, and Dr. Qisheng Zhang, for providing me with honest and thought-provoking suggestions regarding my research study.

Finally, I am so grateful for the former and current labmates of the Liu laboratory. I express thanks to Dr. Ding Xu for teaching me the expression, purification and mutational

analysis of the heparan sulfate sulfotransferases. I would also like to thank former labmates, Dr. Ronald Copeland, Dr. Miao Chen, Dr. Tanya Burch, and Lan Yu, as well as current labmates Dr. Renpeng Liu, Liz Pempe, Dr. Yongmei Xu, Dr. Kai Li, Justin Roberts, Ryan Bullis, Dr. Xianxuan Zhou, Truong Pham, and Dr. Juzheng Sheng for their assistance and expertise during my time in the lab. I would like to send my sincerest gratitude to Courtney Jones, Dr. Renpeng Liu, Liz Pempe, and Sherket Peterson for proofreading and providing me with helpful suggestions to improve my dissertation. Special thanks to the "Liu ladies", Sherket Peterson and Courtney Jones, who continuously offered their advice and support both in and outside the laboratory. Without the help and support of everyone involved, completion of this dissertation would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

ΔUA	$\Delta^{4,5}$-unsaturated uronic acid
ΔUA2S	$\Delta^{4,5}$-unsaturated 2-*O*-sulfated uronic acid
2OST	2-*O*-sulfotransferase
3OST	3-*O*-sulfotransferase
6OST	6-*O*-sulfotransferase
AnMan	2,5-anhydromannitol
AT	Antithrombin
CDSNS	Completely de-*O*-sulfated *N*-sulfated
C_5-epi	C_5-epimerase
CHO	Chinese hamster ovary
CS	Chondroitin sulfate
EST	Estrogen sulfotransferase
EXT	Exostosin
EXTL	Exostosin like
FGF(R)	Fibroblast growth factor (receptor)
GAG	Glycosaminoglycan
Gal	Galactose
GalT-I/II	Galactosyltransferase I/II
GlcNAc	*N*-acetylated glucosamine
GlcNAc6S	*N*-acetylated, 6-*O*-sulfated glucosamine
GlcNAcT-I/II	Glucosaminyltransferase I/II
$GlcNH_2$	Unsubstituted glucosamine

GlcNS	*N*-sulfated glucosamine
GlcUA	Glucuronic acid
GlcUAT-I/II	Glucuronyltransferase I/II
GPI	Glycosylphosphatidylinositol
HIV	Human immunodeficiency virus
HS	Heparan sulfate
HSV-1	Herpes simplex virus-1
IdoUA	Iduronic acid
IdoUA2S	2-*O*-sulfated iduronic acid
LMW HP	Low molecular weight heparin
ITC	Isothermal titration calorimetry
NDST	*N*-deacetylase/*N*-sulfotransferase
NST	*N*-sulfotransferase
PAP	3'-phosphoadenosine 5'-phosphate
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PI-88	Phosphomannopentose sulfate
PSGL-1	P-selectin glycoprotein ligand-1
rFIX	Recombinant factor IX
RPIP-HPLC	Reverse phase ion-pairing high performance liquid chromatography
TPST	Tyrosylprotein sulfotransferase
UDP	Uridine diphosphate
VEGF	Vascular endothelial growth factor
Xyl	Xylose

XylT-I/II	Xylosyltransferase I/II

Chapter I

Introduction

Heparan Sulfate Proteoglycans

Heparan sulfate (HS), a highly sulfated and ubiquitous polysaccharide, is a member of the glycosaminoglycan (GAG) family. Other members of the GAG family include chondroitin sulfate (CS), keratan sulfate, and hyaluronic acid (1). The GAG macromolecules differ in the structure of repeating disaccharide units as well as the level of sulfation on the polysaccharide. Hyaluronic acid is in sharp contrast with the other GAG members in that it possesses no sulfo groups. With the exception of hyaluronic acid, GAGs are covalently attached to a core protein and termed proteoglycans (2). The core proteins of HS proteoglycans consist of syndecans, serglycin, glypicans, agrin, perlecans and collagen XVIII, ranging in size from 32 to 500 kDa (Figure 1) (3). The core proteins can be categorized into two groups, membrane-bound and non-membrane bound. The syndecans and glypicans are associated with the membrane-bound group, containing different types of anchors. Syndecans are anchored to the cell membrane via the transmembrane domain whereas glypicans rely on the glycosylphosphatidylinositol (GPI) anchor (2,4). The non-membrane bound members include perlecan, agrin, and collagen XVIII (2,3). The heparin proteoglycan, serglycin, is a specialized proteoglycan that is an exclusive product of mast cells (5). HS proteoglycans are involved in many biological processes including blood coagulation, wound healing, viral infections, embryonic development, and cancer (4). The

heparan sulfate side chain controls the biological functions through interaction with other biologically relevant proteins whereas the core protein determines the location of the HS within tissues (2).

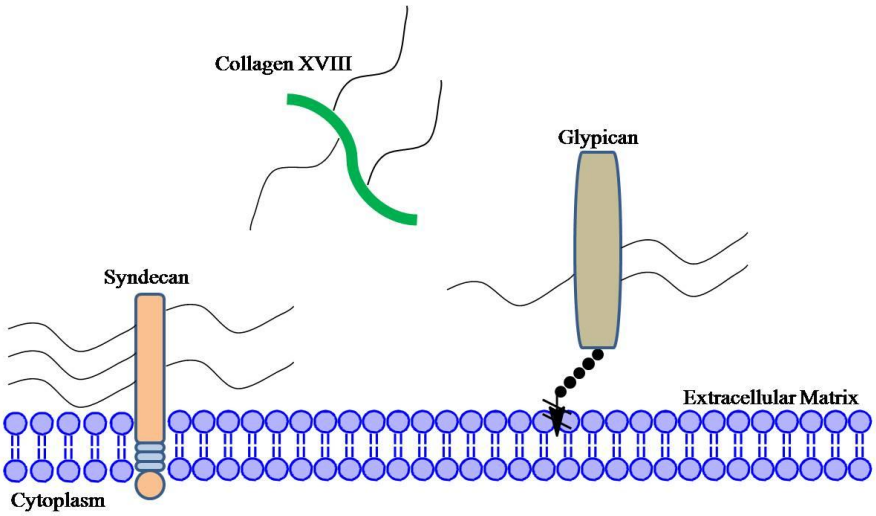


Figure 1. Schematic Representation of the heparan sulfate proteoglycans. Syndecans are integrated into the cell membrane via a transmembrane anchor and carry more than one HS side chain. Glypicans are attached to the membrane via a GPI anchor. Collagen XVIII, perlecan, and agrin are not attached to the cell membrane but instead exist in the ECM.

Chemical Structure of Heparan Sulfate and Heparin

Heparan sulfate is initially synthesized in the *Golgi* apparatus as a repeating structure of D-glucuronic acid (GlcUA) and D-*N*-acetyl glucosamine (GlcNAc) attached through alternating β-1,4 and α-1,4 glycosidic linkages (Figure 2). The disaccharide repeat is synthesized a total of 50-100 times to produce the polysaccharide backbone (6). Heparin, a commonly used anticoagulant, is a structural analog of heparan sulfate. The nascent backbone of both heparin and HS is susceptible to modifications by several biosynthetic enzymes. The glucuronic acid can be converted to L-iduronic acid (IdoUA) by altering the C_5-position, and the uronic acid (GlcUA or IdoUA) can be sulfated at the 2-OH position. For

the glucosamine residue, it can be sulfated at the 3-OH and 6-OH positions and the *N*-position can exist in unsubstituted, acetylated, or sulfated forms. The reactions catalyzed by these biosynthetic enzymes do not proceed to completion thus resulting in heparin/HS with heterogeneous structures. Most of the residues within HS are *N*-sulfated and *N*-acetylated with approximately 1-7% being unsubstituted glucosamine (4,7). Since most of the modification enzymes rely on adjacent *N*-sulfo glucosamine residues, there are several highly sulfated regions within the HS polysaccharide chain known as NS-domains that resemble heparin. The NS domains alternate with larger unmodified regions consisting of mostly GlcNAc residues known as the NA domain. The area between the NS and NA domains contains mixed sequences of alternating *N*-acetylated and *N*-sulfated glucosamine disaccharides termed the NA/NS domain (8,9). The structural diversity of HS has been implicated in several biological functions.

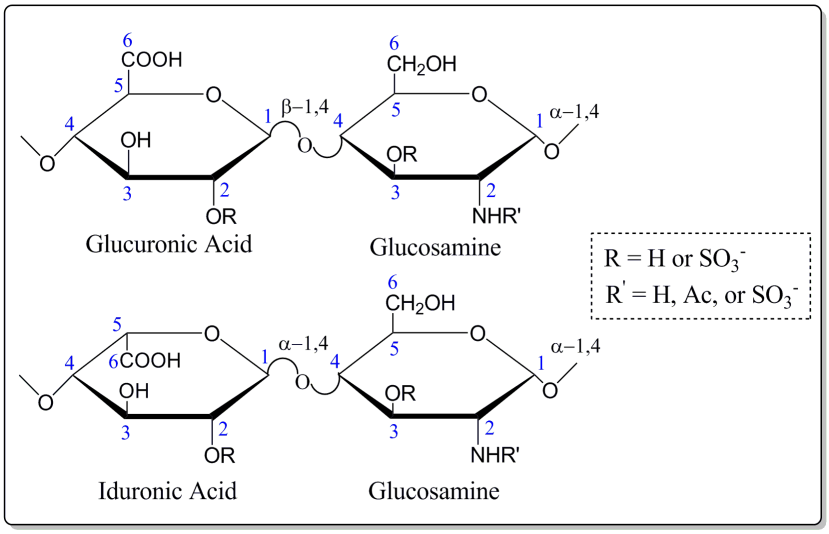


Figure 2. Chemical structure of the disaccharide repeating units of heparan sulfate and heparin. Sulfation (R= $-SO_3^-$) at C6 of glucosamine is common. Sulfation at C2 of iduronic acid is common. Sulfation at C3 of glucosamine is rare. Both *N*-acetyl (R'=acetyl, GlcNAc) and *N*-sulfo glucosamine (R'= $-SO_3^-$, GlcNS) are common. *N*-unsubstituted glucosamine (R' = -H, $GlcNH_2$) is a low abundant component.

The completion of NMR and molecular modeling studies has indicated that heparin and HS adopt an overall helical structure with the sulfate and carboxylate groups projected outward (10). The glycosidic linkages appear to be stiff. The presence of an IdoUA within the chain possesses high flexibility with a conformational equilibrium between chair and skew boat conformations. The presence of sulfate groups affects the conformational equilibrium of IdoUA more so than the overall shape of heparin or HS (11). The overall presentation of negative charge on the exterior of heparin and heparan sulfate allows for extensive ionic interactions with proteins carrying positive charges such as fibroblast growth factors and antithrombin (12-15).

Heparan Sulfate versus Heparin

Both heparin and HS are expressed as proteoglycans, however, they differ in their cellular localization. Heparan sulfate proteoglycans are ubiquitously expressed on the surface and within the extracellular matrix of animal cells, but heparin proteoglycans are exclusive to mast cells (16). Although heparin and HS possess similar structures, they are strikingly different in the level of sulfation and epimerization (15). On average, the heparin disaccharide contains 2.6 sulfo groups per disaccharide whereas HS contains approximately 0.6. In terms of epimerization level, approximately 20% of uronic acid residues within HS are IdoUA whereas heparin contains approximately 90% IdoUA (15,16). With a higher level of sulfation and IdoUA content, heparin can be considered a specialized form of heparan sulfate. The most prevalent disaccharide structure in heparin, constituting 75-90% of disaccharide sequences, is IdoUA2S-GlcNS6S whereas 10-50% of HS disaccharide sequences consist of GlcUA-GlcNAc (16). Therefore, HS has a higher level of structural variability than heparin, possibly explaining its roles in a wide range of biological functions.

Biosynthesis of Heparin and Heparan Sulfate

The biosynthesis of heparin and heparan sulfate takes place in the *Golgi* apparatus, involving a series of specialized glycosyltransferases, sulfotransferases and an epimerase. The HS biosynthetic pathway involves a non-template, enzyme driven process that can be separated into three steps: (1) chain initiation, (2) chain elongation, and (3) chain modification. Essentially all of the biosynthetic enzymes have been successfully cloned and purified to date, enhancing our ability to understand their role in the synthesis of functionally specific heparan sulfate (17).

Chain Initiation

The biosynthesis of the heparan sulfate chain is initiated by the addition of a tetrasaccharide linkage region, -GlcUAβ1-3Galβ1-3Galβ1-4Xylβ1-*O*-[Ser], to specific serine residues on the translated core protein (Figure 3). This tetrasaccharide linkage region, common to heparin, heparan sulfate, dermatan sulfate, and chondroitin sulfate, serves as a precursor to heparan sulfate biosynthesis (18). The addition of each sugar within the linkage region from the activated UDP sugars is performed sequentially by four glycosyltransferases that include xylosyltransferase (XylT), galactosyltransferase I (GalTI), galactosyltransferase II (GalT-II), and glucuronyltransferase I (GlcUAT-I) (Figure 4).

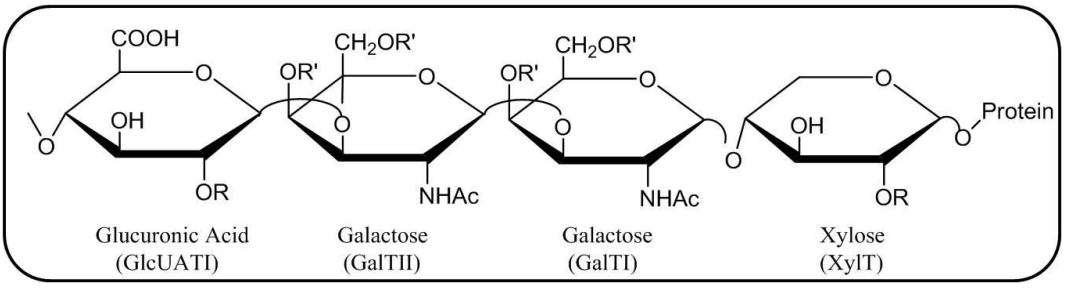


Figure 3. Biosynthesis of the tetrasaccharide linkage region. The tetrasaccharide linkage region (GlcUA-Gal-Gal-Xyl) is attached to specific serine residues on the core protein. The glycosyltransferase responsible for the addition of each sugar is labeled in parentheses under the corresponding sugar unit. R = proton or sulfate; R' = proton or phosphate.

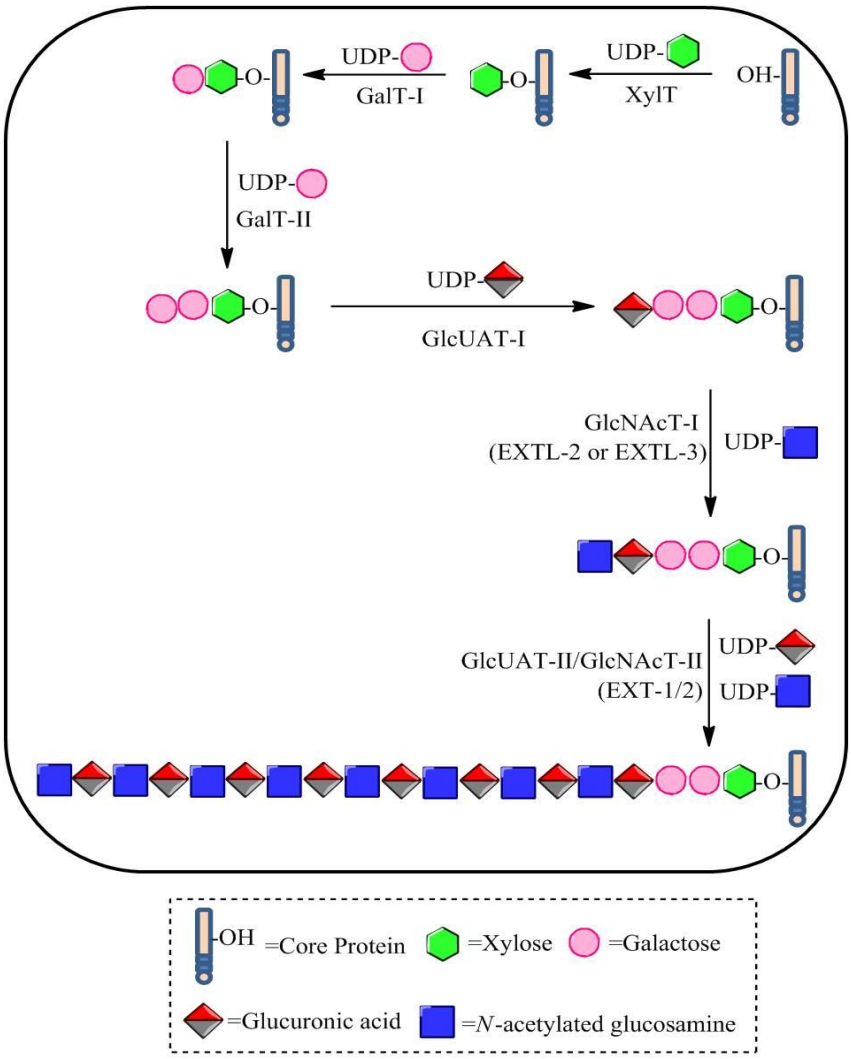


Figure 4. Initiation and elongation of the heparan sulfate nascent chain. Chain initiation begins with the synthesis of the tetrasaccharide linkage region, which is covalently attached to specific serine residues of the core protein. Upon completion of the linkage region, the elongation step begins to complete synthesis of the nascent HS chain.

Xylosyltransferase (XylT)

XylT is responsible for the transfer of xylose from UDP-xylose to specific serine residues of a translated core protein (19-21). This is the first and rate-limiting step in tetrasaccharide linkage biosynthesis (22). Two isoforms of XylT (XylT-I and XylT-II) have been identified in vertebrates with approximately 55% sequence identity among human

XylTs (22,23). The full-length human XylT-I was successfully expressed in mammalian cells in its active form, demonstrating localization in the early *Golgi* cisternae (22). The *N*-terminal 214 amino acid residues of XylT-I were determined to be important for localization to the *Golgi* but not for enzymatic activity (22). Studies have shown that CHO cells deficient in XylT-I expression were unable to synthesize heparan sulfate or chondroitin sulfate, suggesting the importance for XylT-I in initiating synthesis of the linkage region (24),

Until recently, the enzymatic activity of XylT-II had not been elucidated. Using *Pichia pastoris* as a host, an enzymatically active, soluble form of XylT-II was successfully expressed. Studies using various xylosidases to cleave different substrates modified by XylT-II determined that XylT-II is indeed a β-xylosyltransferase (25). As with XylT-I, the postulated concensus sequence for xylosylation is a-a-a-a-G-S-G-a-a/G-a (a represents Asp or Glu), which is consistent with the short minimal motifs G-S-G and G-S-x-G (x represents any amino acid) (26-30). *In vitro* studies with various peptides demonstrated no significant difference in specificity between the two isoforms, suggesting that more specific acceptor molecules are necessary for distinguishing the two enzymes (30).

Galactosyltransferases I and II (GalT-I & GalT-II)

Once the xylose residue has been covalently attached to the serine residue of a core protein, two galactose residues are added by way of two galactosyltransferases, relying on UDP-galactose as the sugar donor. The first galactose residue is added by galactosyltransferase I (GalT-I) while the second galactose is added by galactosyltransferase II (GalT-II). Localized in the medial *Golgi* cisternae, Gal-TI belongs to the β1,4-galactosyltransferase gene family (31) while GalT-II is the sixth member of the β1,3-galactosyltransferase family (32). Transfection of GalTI-deficient CHO cells with GalTI

cloned DNA completely restored both heparan sulfate and chondroitin sulfate biosynthesis (33). The importance of GalT-II *in vivo* was confirmed by the noticeable blocking of GAG biosynthesis upon silencing of GalT-II in HeLa S3 cells (32). These results demonstrate the functional importance of both GalT-I and GalT-II to heparan sulfate biosynthesis. Mutations of GalT-I have been implicated in the onset of a progeroid variant of Ehlers-Danlos syndrome, a connective tissue disorder due to little proteoglycan biosynthesis in humans (34,35). However, more studies must be completed in order to determine if other glycosyltransferases such as XylT, GalT-II, or GlcUAT-I play a role in this disorder.

Glucuronyltransferase I (GlcUAT-I)

The tetrasaccharide linkage region synthesis is completed by the addition of the final residue, a glucuronic acid. The transfer of glucuronic acid from the sugar donor UDP-glucuronic acid to the second galactose residue via a β1,3-linkage is catalyzed by glucuronyltransferase I (GlcUAT-I) . After several attempts, the GlcUAT-I was successfully purified to homogeneity from human placenta (36-38). Enzymatic assays determined that modifications on the xylose and galactose residues of the tetrasaccharide linkage region, specifically a 2-*O*-phosphate group on xylose and 6-*O*-sulfo group on the first galactose, enhanced the activity of GlcUAT-I (39). It is important to note that these modifications have not been identified simultaneously on the chain (29) however there is a possibility that modifications to the tetrasaccharide region may play an important role in completing GAG biosynthesis.

The crystal structure of GlcUAT-I has been solved in the presence of donor and acceptor substrates (40,41). A ternary complex of GlcUAT-I with the sugar donor UDP and the acceptor substrate Galβ1,3Galβ1,4Xyl revealed that the enzyme mainly recognizes

Galβ1,3Gal with xylose residing outside the substrate binding cavity, potentially serving as a spacer between the Galβ1,3Gal and the core protein (40). The binary complex of GlcUAT-I with the complete UDP-GlcUA sugar donor provides support for an in-line displacement mechanism for this glycosyltransferase with a glutamic acid implicated as the potential catalytic base (41).

With approximately 43% sequence identity, an additional β1,3-glucuronyltransferase GlcUAT-P, known to add GlcUA to glycoproteins, was determined to possess similar substrate specificity to GlcUAT-I (42). Interestingly, an *in vivo* study utilizing GlcUAT-I deficient CHO cells demonstrated restoration of GAG synthesis upon transfection with either GlcUAT-I or GlcUAT-P (42). Despite their similar substrate specificities, GlcUAT-I and GlcUAT-P have different tissue expression, with GlcUAT-I being expressed in all examined tissues and GlcUAT-P exclusively expressed in the brain (42). The exclusive expression of GlcUAT-P suggests a potential role for GAG production in neural tissues.

Chain Polymerization

Following preparation of the tetrasaccharide linkage region (GlcUA-Gal-Gal-Xyl), the heparan sulfate nascent chain is biosynthesized by the addition of the first *N*-acetyl glucosamine residue to the non-reducing end GlcUA by CHO derived GlcNAc transferase I (GlcNAcT-I). The addition of the initial GlcNAc residue to the tetrasaccharide linkage region is a commitment step for heparan sulfation polymerization. The GlcNAcT-I enzyme is not to be confused with the GlcNAc transferase II (GlcNAcT-II) that is responsible for the addition of GlcNAc residues during chain polymerization/elongation (43). The GlcNAcT-I activity was identified in two mammalian enzymes encoded by the exostosin-like (EXTL) family, EXTL-2 and EXTL-3 (44,45). EXTL-2 possesses solely GlcNAcT-I activity whereas

EXTL-3 harbors both GlcNAcT-I and GlcNAcT-II activities (45). *In vitro* studies using *N*-acetyl heparosan oligosaccharides and a known synthetic substrate for GlcNAcT-I activity, GlcUAβ1–3Galβ1-*O*-C_2H_4NH-benzyloxycarbonyl, showed EXTL-3 activity with both substrates (45). This result suggests that EXTL-3 could be involved in both initiation and elongation. *In vivo* studies are essential for understanding the role of EXTL-2 and EXTL-3 in initiating HS polymerization. Mutations in the zebrafish EXT-like *boxer* gene, homologous to human EXTL-3, resulted in defective fin and branchial arch development, possibly due to defective FGF10 signaling (46). The zebrafish mutant presented significantly reduced HS levels, further supporting the role of EXTL-3 in HS initiation (47). A GlcNAc transferase known as rib-2 was identified in *C. elegans* and determined to be involved in both heparan sulfate chain initiation and elongation with similar substrate specificity to EXTL-3, suggesting a distinct mechanism from mammals (48).

Once committed to HS biosynthesis with the addition of the first GlcNAc residue by GlcNAcT-I, chain elongation begins by the alternating addition of GlcUAβ4 and GlcNAcα4 residues, from UDP-GlcUA and UDP-GlcNAc respectively, by a set of mammalian glycosyltransferases belonging to the exostosin (EXT) gene family known as EXT-1 and EXT-2. The EXT family was initially recognized as causative genes of a genetic bone disorder as well as tumor suppressor genes (49,50). Mutations in EXT-1 or EXT-2 are also responsible for multiple hereditary exostoses, hallmarked by tumors within the cartilage (51,52). When expressed, EXT-1 and EXT-2 were determined to exhibit both GlcUA and GlcNAc transferase activities, indicating their functional role as a co-polymerase (53,54). Studies have demonstrated that EXT-1 and EXT-2 form a stable complex accumulating in the *Golgi* apparatus with significantly higher glycosyltransferase activities when co-

expressed than when expressed individually (52,54). A cell line defective in EXT-1 expression demonstrated no compensation by EXT-2 in the absence of EXT-1, suggesting that the EXT-1/EXT-2 complex represents the active form within cells (52). Mice deficient in EXT-1 exhibited early embryonic lethality as well as poor gastrulation and defective bone growth (55). Like with EXT-1 disruption, the EXT-2 knockout mice show embryonic lethality and poor gastrulation (56). Targeted disruption of either the EXT-1 or EXT-2 genes in mice resulted in abolition of HS synthesis (55,56). Kim *et al.* reported *in vitro* polymerization of the synthetic linkage region of GlcUA-Gal-*O*-C_2H_4NH-benzyloxycarbonyl upon co-transfection with EXT-1 and EXT-2 (50). Interestingly, no polymerization occurred when a mixture of individually expressed EXT-1 and EXT-2 were incubated with the natural tetrasaccharide linkage region of GlcUAβ1-3Galβ1-3Galβ1-4Xylβ1-*O*-Ser (50), suggesting the importance of hydrophobic residues near the linkage region that can be achieved by the artificial aglycon, C_2H_4NH-benzyloxycarbonyl (45).

In addition to EXT-1 and EXT-2, there are two other mammalian enzymes, EXTL1 and EXTL3, which are known to possess GlcNAcT-II activity (45). Neither EXTL1 nor EXTL3 demonstrated GlcUAT-II activity as determined by *in vitro* studies using *N*-acetyl heparosan oligosaccharides with the non-reducing terminal GlcNAc (45). As for EXT-1 and EXT-2, there was no GlcNAc transfer to the natural tetrasaccharide linkage for EXTL1 or EXTL2 further supporting the potential necessity for hydrophobic residues near the linkage region like with the artificial aglycon on the synthetic substrate (45). The role of EXTL1 and EXTL2, both possessing GlcNAcT-II activity, in HS chain polymerization remains to be determined. Further studies are necessary to determine whether EXTL1 and/or EXTL3 interact with EXT-1 and/or EXT-2 to complete polymerization.

Chain Modification

Once the native heparan sulfate polysaccharide backbone has been synthesized, it is then susceptible to numerous modifications involving a series of specialized enzymes including an epimerase and various sulfotransferases as illustrated in Figure 5 (57-59). The HS backbone is first modified by *N*-Deacetylase/*N*-Sulfotransferase (NDST) which serves to convert *N*-acetylated glucosamine (GlcNAc) residues to *N*-sulfo glucosamine (GlcNS). Following *N*-sulfation, C_5-epimerase converts D-glucuronic acid (GlcUA) to L-iduronic acid (IdoUA). The backbone is then modified by several *O*-sulfotransferases including 2-*O*-sulfotransferase (2OST) that introduces 2-*O*-sulfo groups to GlcUA or IdoUA units as well as 6-*O*-sulfotransferase (6OST) and 3-*O*-sulfotransferase (3OST) which add 6-*O*-sulfo and 3-*O*-sulfo groups to the glucosamine unit respectively. The sulfotransferases rely on a biologically active sulfo donor, 3'-phosphoadenosine 5'-phosphosulfate (PAPS) that is synthesized by PAPS synthase in the cytoplasm and nucleus of animals (Figure 6A) (60). Once synthesized, PAPS must be transported to the *Golgi* apparatus by the action of the PAPS transporter for heparan sulfate biosynthesis (61,62).

Combined structural and mutational analyses have provided some insight into the catalytic mechanism of the HS sulfotransferases especially that of 3OST-3 whose crystal structure includes a ternary complex of human 3OST-3 with PAP and a tetrasaccharide HS substrate (63). The ternary complex of 3OST-3 suggests that sulfo group transfer occurs by a S_N2-like in-line displacement mechanism with the potential catalytic base, such as glutamic acid, deprotonating the 3-OH position of glucosamine for nucleophilic attack on the sulfo group of PAPS as illustrated in Figure 6B. These chain modification reactions do not proceed to completion, resulting in high structural variability of the HS products (64).

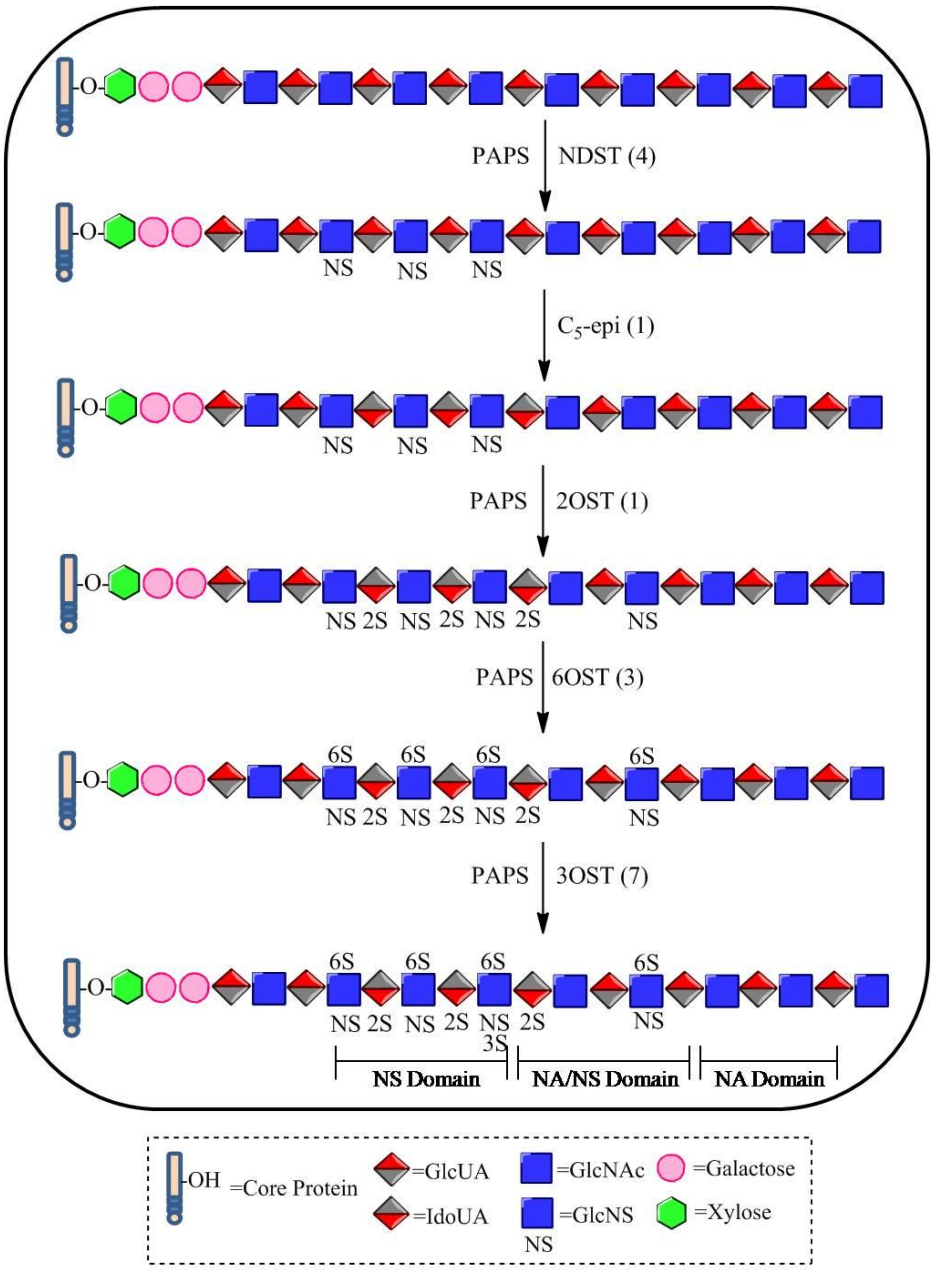


Figure 5. Modification of the heparan sulfate polysaccharide backbone. Once the HS nascent polymer is synthesized by specialized glycosyltransferases, the chain is susceptible to several enzymatic modifications including *N*-sulfation and epimerization of GlcUA to IdoUA followed by *O*-sulfation by different HS *O*-sulfotransferases. The sulfotransferase reactions utilize the universal sulfo donor PAPS to transfer the sulfo groups to the chain. Several of the modification enzymes exist in more than one isoform as indicated by the number in parentheses. *N*-sulfation is a prerequisite for subsequent modifications but does not proceed to completion, resulting in three domains differing in the level of *N*-sulfation.

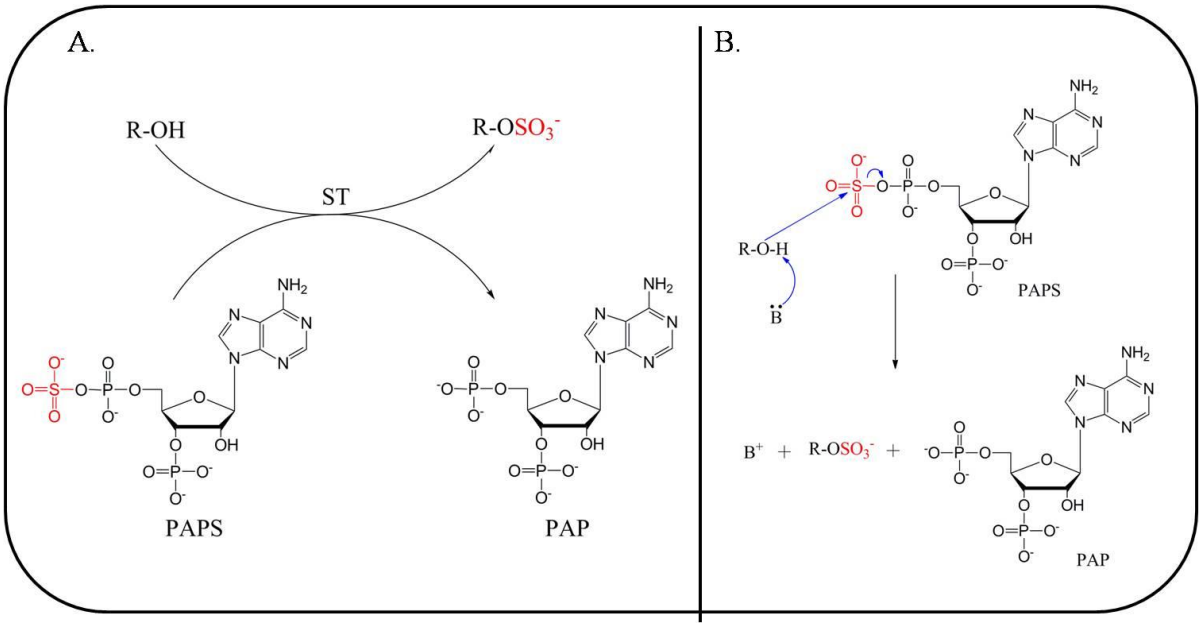


Figure 6. Mechanism of the heparan sulfate sulfotransferases. **(A)** General mechanism of HS sulfotransferases. The sulfotransferase catalyzes the transfer of the 5'-sulfo group from the donor PAPS to the heparan sulfate substrate resulting in the sulfated product and PAP. **(B)** Proposed catalytic mechanism of HS sulfotransferases.

Factors Controlling Substrate Specificity

The biosynthesis of HS is not a template-driven process therefore the final HS products are dictated by the substrate specificity of the biosynthetic enzymes. The substrate recognition of these enzymes is controlled at the monosaccharide level. This control is dependent upon the regioselectivity of the residue at the site of modification (2). For example, 2OST recognizes either GlcUA or IdoUA and specifically transfers a sulfo group to the 2-OH position to generate GlcUA2S or IdoUA2S respectively. The specificity of the enzymes is also controlled at the oligosaccharide level through recognition of the sulfation pattern surrounding the modification site. With 2OST, a non-reducing end GlcNS residue is required for sulfation (65). The lack of homogeneous polysaccharide substrates has hindered our ability to complete elaborate substrate specificity studies for the biosynthetic enzymes.

With the exception of C_5-epi and 2OST that exist in a single isoform, the presence of multiple isoforms of the modification enzymes including NDST, 6OST, and 3OST serves to regulate the sulfation pattern of the final heparan sulfate structures (2,66-69)**.** Studies have indicated that the different isoforms recognize different sulfation patterns around the modification site, implicating their role in preparing specific sequences within the HS chain (2,18). The presence of multiple isoforms of the modification enzymes suggests that the biosynthesis of heparan sulfate is strictly regulated in vertebrates.

The substrate specificity of HS sulfotransferases controls the sulfation patterns of HS, permitting HS to exhibit a specific function, however, limited knowledge regarding the mechanism of these enzymes has hindered our ability to prepare functionally-specific HS. Therefore, understanding the substrate recognition mechanism used by the sulfotransferases is necessary for advancing the enzymatic synthesis of HS as well as delineating the biosynthetic mechanism of HS. Coupling a structural biology approach with site-directed mutagenesis offers an effective methodology for understanding the substrate recognition mechanism of the HS sulfotransferases. The crystal structures of several HS sulfotransferases, including *N*-sulfotransferase domain of NDST and three isoforms of 3OST, have already been elucidated. However, more extensive studies must be completed of the structurally unknown 2OST and 6OST to gain more insight into their molecular mechanism.

Glucosaminyl N-Deacetylase/N-Sulfotransferase (NDST)

The bi-functional enzyme, *N*-Deacetylase/*N*-Sulfotransferase, is the initial step in the modification of both heparan sulfate and heparin, which is absolutely mandatory for subsequent epimerization and *O*-sulfations to occur (67). The NDST enzyme catalyzes

deacetylation of the GlcNAc residue to produce free amine at the C-2 position, immediately followed by *N*-sulfation to produce GlcNS (70) (Figure 7).

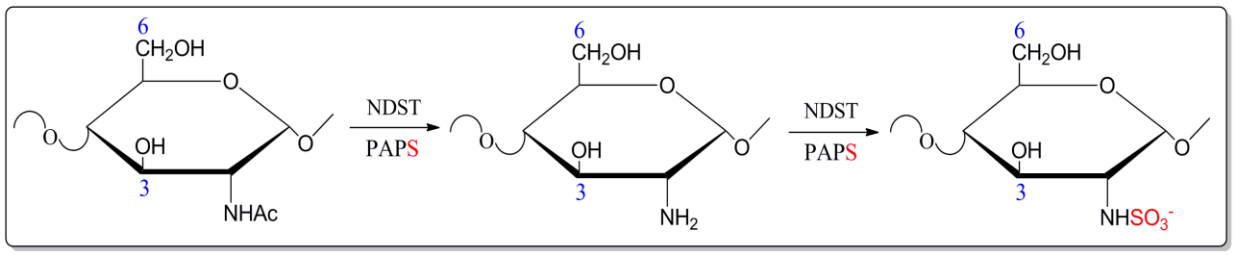


Figure 7. Reactions catalyzed by heparan sulfate NDST. NDST catalyzes deacetylation of glucosamine (GlcNAc) followed by *N*-sulfation in the presence of the sulfo donor PAPS to produce *N*-sulfo glucosamine (GlcNS).

Four isoforms of NDST have been identified in vertebrates including both human and mouse. These four isoforms have been successfully cloned and possess distinct tissue expression patterns (71-76). NDST-1 and NDST-2 being uniformly expressed across all examined tissues whereas NDST-3 and NDST-4 were predominately expressed in the brain (71). There are two subtypes of NDSTs, consisting of NDST-1/2 and NDST-3/4 (71). NDST-1 and NDST-2 share 70% sequence identity whereas NDST-3 and NDST-4 share 80% identity. NDST-2 is highly expressed in mast cells and has been speculated to be involved in biosynthesis of heparin but not heparan sulfate (74). The isoforms possess strikingly different enzymatic properties. NDST-1 and -2 demonstrated similar levels in deacetylase and sulfotransferase activities but with NDST-1 having a slightly higher level of sulfotransferase activity (71). On the other hand, NDST-3 possesses high deacetylase activity but weak sulfotransferase activity whereas NDST-4 presented weak deacetylase activity and high sulfotransferase activity (71). The distinct activity and expression patterns of the four isoforms suggest a potential for varying substrate recognition among each

isoform. Unfortunately, substrate specificity studies have yet to be completed for the NDSTs therefore the specific structure recognized by this enzyme is currently unknown (2).

The NDST protein, comprised of 883 amino acid residues with an approximate molecular weight of 100 kDa (74), is divided into two domains: the *N*-terminal deacetylase domain and the *C*-terminal sulfotransferase domain. The deacetylase domain of NDST comprises the first 600 amino acid residues whereas the sulfotransferase domain comprises the final 283 residues (77). Both domains have been expressed individually and determined to exhibit their anticipated activities (78,79). A crystal structure of the *C*-terminal NST domain of human NDST-1 in complex with the inactive sulfo donor, PAP, was determined at a resolution of 2.3 Å (79) (Figure 8). The overall fold of NST-1 is consistent with that of a type II membrane-bound sulfotransferase with a α/β fold containing a central five-stranded parallel β-sheet and a large open cleft with a hydrophilic surface for HS binding (79). The substrate binding site of NST-1 is very distinct from the structurally known cytosolic sulfotransferases such as estrogen sulfotransferase (EST), which possesses a hydrophobic binding pocket for estradiol (80). The catalytic mechanism of NST was proposed based on a molecular model of NST-1 with a docked hexasaccharide in the active site (81). Analysis of residues within the active site revealed a potential role for Glu-642 as the catalytic base due to its hydrogen bonding distance with the NH_3^+ group of glucosamine and complete abolition of sulfotransferase activity when mutated (81). Although the crystal structure of the *N*-sulfotransferase domain of NDST-1 has provided insight into its catalytic mechanism, the crystal structure of the *N*-deacetylase domain has yet to be determined, limiting our knowledge of the deacetylation mechanism. A better understanding of the deacetylase

mechanism of NDST could help explain the difference in deacetylase activities as well as decipher the substrate requirements among the isoforms.

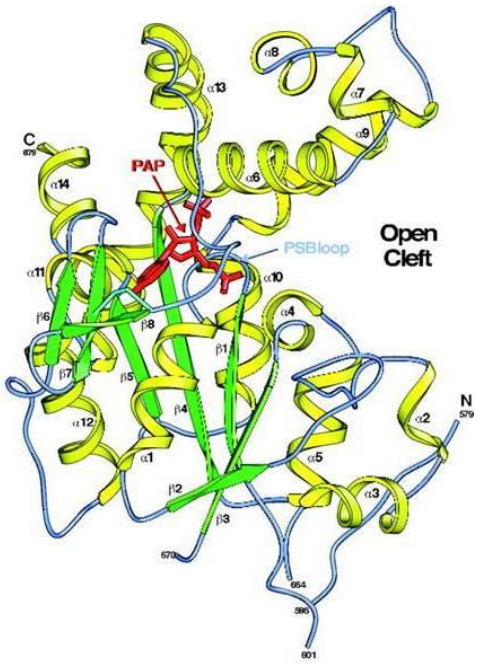


Figure 8. Ribbon representation of the structure of the *N*-sulfotransferase domain of human NDST-1 in complex with PAP. Helices are colored yellow, β-strands in green, random coil in blue, and PAP in red. Structure taken from (79).

The role of *N*-sulfation in mammalian development has been established through several NDST knockout mouse models. Mice deficient in NDST-1 exhibited incomplete lung development and respiratory distress, resulting in neonatal death (82). In addition to lung formation, NDST-1 expression is essential for forebrain, eye and face development (83). Recently, a prominent role for *N*-sulfation in inflammation was established through a knockout study of NDST-1 within endothelial cells and leukocytes of mice. The absence of NDST-1 expression within these specific tissues resulted in a decreased inflammatory response, including a reduction in the recruitment of inflammatory cells as well as a decreased allergen response (84).

As mentioned previously, it has been speculated that NDST-2 is exclusive for synthesis of heparin and not heparan sulfate in mast cells (74). It was previously determined that mast cells have high levels of expression of NDST-2 but little to none of NDST-1 (76). Indeed, mice deficient in NDST-2 were incapable of producing sulfated heparin, possessing fewer connective-tissue mast cells as well as significantly reduced histamine and mast-cell proteases (85). Recently, it was determined that NDST-2 plays a key role in the granule storage capacity of mast cells which is promoted by serglycin, a glycoprotein with several attached HS/CS chains. NDST-2 expression increased during mast cell maturation, suggesting a role in formation of highly sulfated heparin attached to the serglycin core protein (86). NDST-2 is essential for heparin biosynthesis however the role in synthesizing heparan sulfate is less clear since there was no apparent difference in the level of *N*-sulfation among wild-type and NDST-$2^{-/-}$ mice (85).

A targeted disruption of NDST-3 in mice presented no significant alteration in development or homeostasis, only demonstrating slight abnormalities in hematological parameters and behavior (87). A series of double knockout mice studies were completed to determine a potential compensatory effect by the other NDSTs when NDST-3 expression was eliminated. A NDST-2/NDST-3 deficient mouse appeared normal whereas mice deficient in both NDST-1 and NDST-3 did not survive birth and exhibited developmental defects caused by severe HS undersulfation, suggesting a potential role for NDST-3 in HS biosynthesis in the absence of NDST-1 (87). Any potential roles for NDST-4 have yet to be established.

Glucuronosyl C_5-Epimerase (C_5-epi)

Heparan sulfate C_5-epimerase (C_5-epi) is the first modification known to occur on the uronic acid residue, catalyzing the conversion from D-glucuronic acid to L-iduronic acid through abstraction and readdition of a proton at the C_5 position of the uronic acid residue by way of a carbanion intermediate (88,89) (Figure 9). Because of the reversibility of the reaction catalyzed by C_5-epi, the resultant polysaccharide contains both GlcUA and IdoUA units. The conversion of GlcUA to IdoUA increases the flexibility within HS since the IdoUA unit can adopt several conformations including the low energy chair ($^4C_1$ and $^1C_4$) and skew boat ($^2S_0$) conformations (11,90). The GlcUA, on the other hand, only exits in a single $^4C_1$ conformation (11). Studies have indicated that the flexibility of the HS caused by the presence of IdoUA contributes to the various biological functions (14).

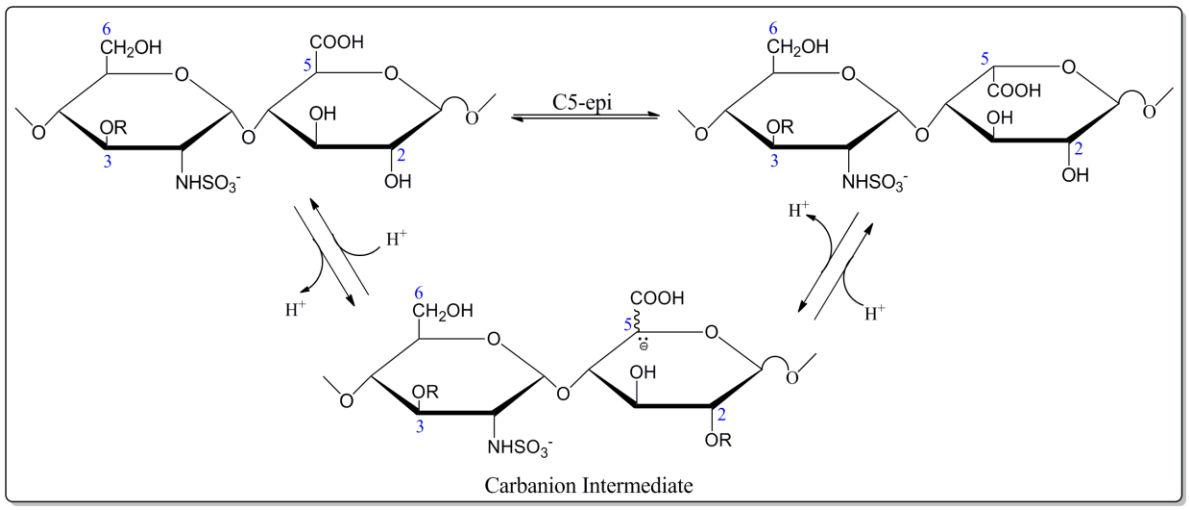


Figure 9. Proposed reaction mechanism of C_5-epi. The C_5-epi catalyzes the abstraction of a proton from the C-5 position of the uronic acid residue to form a carbanion intermediate. The readdition of a proton results in conversion from the D-glucuronic acid to L-iduronic acid. The epimerase reaction recognizes an *N*-sulfo glucosamine residue at the non-reducing end.

C_5-epi consists of 618 amino acids with a molecular weight of approximately 70 kDa (91). The C_5-epi was successfully purified from mouse mastocytoma (91,92) and bovine liver (93) and cloned from bovine lung (89). The enzyme exists in a single isoform among all vertebrate and invertebrate species with human C_5-epi possessing greater than 95% sequence identity to mice (91). Previous substrate specificity studies have indicated that a non-reducing end *N*-sulfo glucosamine (GlcNS) residue is required for the epimerization reaction to occur (94). Therefore, -GlcNS-GlcUA-GlcNAc- is an appropriate substrate for C_5-epi while -GlcNAc-GlcUA-GlcNS- is not (94). The presence of *O*-sulfo groups have been determined to inhibit the reactivity of the uronic acid residue with C_5-epi, suggesting that epimerization precedes *O*-sulfation within heparan sulfate biosynthesis (94).

In vivo functions of C_5-epi were uncovered through gene knockout experiments. Targeted disruption of C_5-epi in mice resulted in abnormal heparan sulfate structure, lacking iduronic acid residues and possessing a severely distorted sulfation pattern (68). C_5-epi knockout mice are lethal, manifesting immature lung phenotype, abundant skeletal abnormalities and kidney agenesis (95). It should be noted that C_5-epi and 2OST reportedly form a complex *in vivo*, suggesting the intimate relationship between those two enzymes (96). As expected, the phenotypes displayed by C_5-epi-knockout mice are similar to those observed in 2OST-null mice (239).

With the knowledge gained from understanding the catalytic mechanism of C_5-epimerase, Campbell *et al.* reported a method to measure the epimerase activity by determining the released tritium, using a site-specific ^{3}H-labeled polysaccharide substrate. In this method, the ^{3}H-label was introduced at the C_5 position of GlcUA/IdoUA unit of CDSNS heparin (93). The C_5-epi de[^{3}H]protonates at the C_5-position of a GlcUA or IdoUA unit,

dissociating the ^{3}H-label from the polysaccharide. The activity of C_5-epi can be determined by measuring the amount of ^{3}H-label released from the polysaccharide (93). Although this method was successfully employed to purify the C_5-epi (89), the preparation of the site-specifically ^{3}H-labeled polysaccharide involves the use of a large amount of [^{3}H]H_2O, which cannot be readily prepared in a standard academic lab. The development of an efficient method for assaying C_5-epi activity could enhance our understanding of the mechanism of C_5-epi.

Uronosyl 2-O-Sulfotransferase (2OST)

Heparan sulfate 2OST is the only sulfotransferase known to modify the uronic acid residue, catalyzing the transfer of a sulfo group from PAPS to the 2-OH position of uronic acid. Like epimerase, 2-*O*-sulfation only occurs within the *N*-sulfated rich regions of heparan sulfate, with a non-reducing end GlcNS residue being an absolutely requirement (65). Substrate specificity studies have indicated that 2OST can modify both GlcUA and IdoUA but possesses a selective preference for iduronic acid when both residues are present within the HS chain (97,98) (Figure 10).

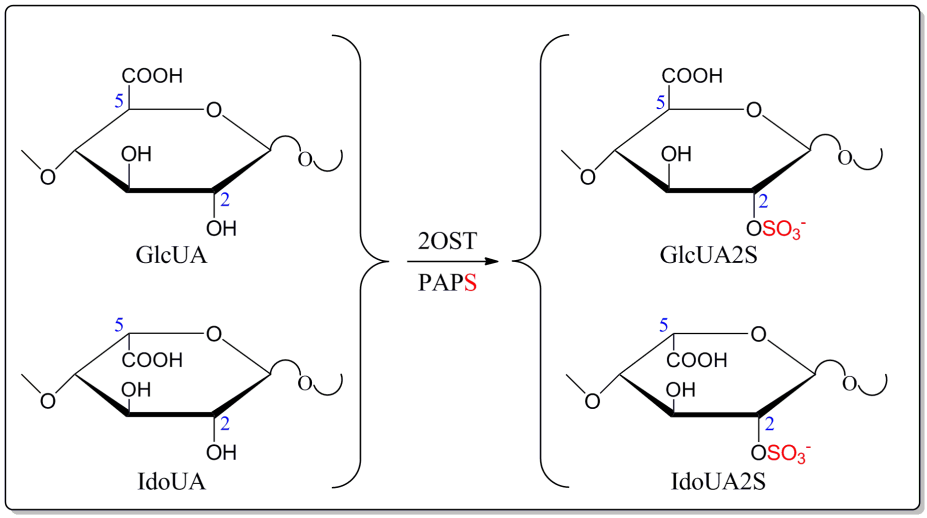


Figure 10. Reactions catalyzed by heparan sulfate 2OST. 2OST sulfates both GlcUA and IdoUA with a selective preference for IdoUA.

The 2OST enzyme, significantly smaller than C_5-epi, consists of 356 amino acid residues with a molecular weight of 43 kDa, originally purified to homogeneity from Chinese hamster ovary cells with gel filtration analysis revealing the protein may exist as a dimer or trimer (99). 2OST is present in a single isoform in most vertebrate and invertebrate genomes, possessing high sequence conservation among species. Human 2OST shares 97% sequence identity with mouse and 92% identity with chicken. The high degree of sequence identity suggests a central physiological role in these organisms. The role of 2-*O*-sulfation in fine structure formation has been examined in several species. The 2OST knockout mice demonstrate renal agenesis and die in the neonatal period, demonstrating retarded eye development and skeletal over-mineralization (100). In *C. elegans,* 2OST has been determined to be essential for axon migration or guidance and nervous system development (101). It was also determined that 2OST is important for proper limb bud formation in chicken (102,103). Dependence on 2-*O*-sulfation is notably less critical in *Drosophila* fibroblast growth factor signaling. It was suggested that 2OST -/- mutant synthesizes HS with a higher level of 6-*O*-sulfation, which may compensate for the reduced 2-*O*-sulfation in *Drosophila.*

The lack of an X-ray crystal structure for 2OST had hindered our understanding of its substrate recognition mechanism. Extensive mutagenesis studies based on homology to other structurally known sulfotransferases (EST and 3OST-1) identified key amino acid residues involved in the binding of 2OST to PAPS and the HS substrate (104). Isothermal titration calorimetry (ITC) was utilized to determine the binding of the 2OST mutants to PAP; PAPS is chemically unstable and therefore cannot be used with ITC. Residues Lys-83, Thr-84, Ser-86, and Thr-87 were speculated to interact through hydrogen bonding with the 5'-

phosphosulfate group of PAPS (104) while Arg-164 and Ser-172 contributed to binding the 3'-phosphate group. Mutation of these residues significantly reduced and often eliminated PAP binding, suggesting their importance in the binding of 2OST to PAPS.

A structural neighbor search using GenTHREADER revealed that 2OST is structurally similar to estrogen sulfotransferase (EST) and 3-*O*-sulfotransferase (3OST) (104), especially within the β7 region of EST known to possess the catalytic base, His-108. The EST catalytic base aligns with His-142 of 2OST, suggesting that His-142 may serve as the base for 2OST. Another histidine residue, His-140, for 2OST may be within a similar distance to the catalytic site, therefore both histidine residues were mutated and presented a similar reduction in sulfotransferase activity compared to the wild-type enzyme. A double mutation of H140A/H142A completely abolished sulfotransferase activity while maintaining PAP binding, suggesting an important role for these two histidine residues to the activity of 2OST (104). Using the ternary complex of 3OST-3 with PAP and a HS tetrasaccharide substrate, several residues of Arg-80, Arg-178, Asp-181, Arg-189, and Arg-288 were identified to potentially influence substrate binding (104). The revelation of a crystal structure for 2OST provided validation to the preliminary mutational analysis as well as heightening our understanding of the mechanism of action of the enzyme.

Glucosaminyl 6-O-Sulfotransferase (6OST)

Heparan sulfate 6OST catalyzes the transfer of a sulfo group to the 6-OH position of the glucosamine residue within the heparan sulfate chain (Figure 11). 6-*O*-sulfation occurs mostly within the NS domain of HS however it can occur within the NA domain to generate GlcNAc6S (66). This is the only sulfation known to occur within the NA domain of heparan sulfate since both 2OST and 3OST modifications rely on *N*-sulfation for the reactions to

occur (18). 6-*O*-sulfation is an essential modification for generating HS structures with specific biological functions, including antithrombin binding for blood anticoagulation as well as fibroblast growth factor (FGF) binding for triggering FGF-mediated signaling pathways (105,106).

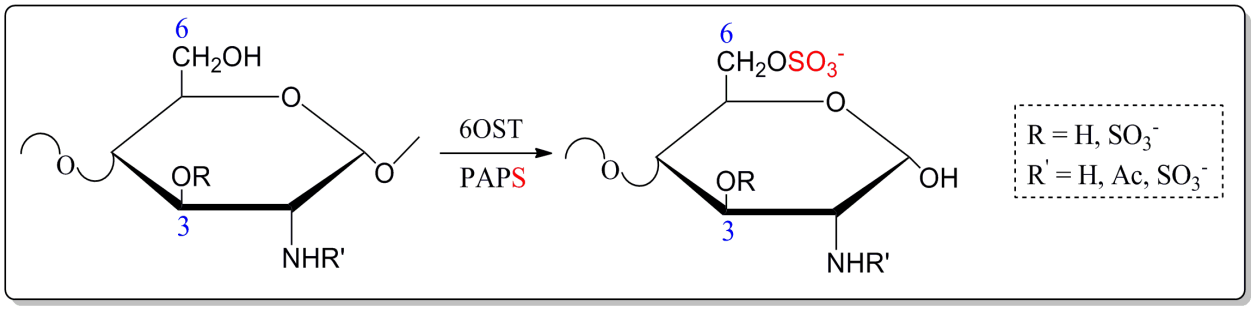


Figure 11. Reactions catalyzed by heparan sulfate 6OST. 6OST can transfer a sulfo group to either a GlcNS or GlcNAc residue to generate GlcNS6S and GlcNAc6S respectively. The enzyme recognizes either GlcUA or IdoUA at the non-reducing end of the glucosamine residue.

The full-length 6OST enzyme contains 401 amino acids with a molecular weight of approximately 47 kDa (107), demonstrating type II transmembrane protein properties. The enzyme was originally isolated to homogeneity from Chinese hamster ovary cells with gel filtration analysis revealing that 6OST exists as a monomer (108). Invertebrates including *Drosophila* and *C. elegans* possess only a single ortholog of 6OST however vertebrates have a more diverse distribution with mouse and human having three isoforms (66, 107, 109-111). With a deletion of 40 amino acids, a shorter form of human 6OST-2, 6OST-2S, due to alternative splicing was identified in humans that maintained similar substrate preference to the longer form 6OST-2 (112). The shorter form of 6OST-2 has been identified in chicken, which is known to possess two isoforms (111,113).

Preliminary substrate specificity studies have been completed for the 6OST isoforms however the results are inconsistent. In 2000, Habuchi *et al.* determined that all of the 6OST

isoforms were capable of sulfating *N*-sulfo glucosamine residues within completely de-*O*-sulfated N-sulfated (CDSNS) heparin and *N*-sulfo heparosan but not *N*-acetylated glucosamine (114). This study suggests that none of the isoforms are capable of recognizing GlcNAc residues. Comparison of the susceptibilities of the 6OSTs to either a GlcUA-containing polysaccharide (*N*-sulfo heparosan) or an IdoUA-containing polysaccharide (CDSNS heparin) demonstrated a slight preference among the isoforms. The 6OST-1 appears to prefer IdoUA-GlcNS sequences whereas 6OST-3 has equal recognition for IdoUA-GlcNS and GlcUA-GlcNS. Interestingly, 6OST-2 appears to depend on substrate concentration with an increase in activity as the amount of GlcUA-GlcNS increased but a decrease in activity as IdoUA-GlcNS increased (114). In 2003, Smeds *et al.* further investigated the effect of the uronic acid residue adjacent to glucosamine as well as the presence of 2-*O*-sulfo groups on the specificity of the 6OST isoforms (115). The results of this study were highly contradictory to the previous report in that all the isoforms were capable of sulfating both GlcNS and GlcNAc residues, suggesting similar specificity among the isoforms. All of the isoforms targeted glucosamine residues to the reducing end of either a GlcUA or IdoUA although IdoUA-containing disaccharides were preferred whenever present. The presence of 2-*O*-sulfo groups did not affect the recognition by the 6OST isoforms (115,116). One major contradiction is that 6OST-1, originally found to prefer IdoUA-GlcNS, demonstrated a preference for GlcUA-GlcNS in this study, which can only be explained by the use of heterogeneous polysaccharide substrates (115). The ambiguity of the specificity for the 6OST isoforms can only be resolved by the use of homogeneous heparan sulfate substrates with defined structures.

The biological and physiological relevance of 6OST has been established through systematic deletion of the 6OST isoforms in different organisms. Targeted disruption of 6OST-1 in mice demonstrated late embryonic lethality, abnormal placentation, and decreased growth, which is strikingly similar to phenotypes seen in Wnt-2 knockout mice (117,118). Wnt-2 is known to initiate expression of vascular endothelial growth factor-A (VEGF-A), an inducer of angiogenesis, and has an affinity for heparan sulfate (119). The 6OST-1 knockout mice exhibit reduced Wnt-2 binding and reduced levels of VEGF-A, suggesting 6OST is important for Wnt-2 recognition to assist with placental vasculogenesis (117). Analysis of the HS structures from different tissues from the 6OST-1 null mice showed a marked decrease in the levels of GlcNAc6S and GlcNS6S in the liver, kidney and lungs, which were previously identified to express high levels of 6OST-1 (114). This marked decrease implies that 6OST-1 may be important for the development of these specific tissues. A morpholino knockdown of 6OST in zebrafish, possessing greater than 50% identity with human 6OSTs, demonstrated a role for 6-*O*-sulfation in muscle development that appears to regulate the Wnt-signaling pathway (120). Taken together, these knockout studies support the importance of 6OST for developmental processes in vertebrates.

The role of 6-*O*-sulfation has also been analyzed in simpler organisms, including *Drosophila* and *C. elegans*. Disruption of 6OST expression in *Drosophila* resulted in embryonic lethality and defective tracheal development (109). The *Drosophila* tracheal system shares several features in common with the vertebrate vascular system, further implicating a role of 6OST in vascular development as is indicated by the mouse model (121). This study provides strong *in vivo* evidence that 6-*O*-sulfated HS is crucial for FGF signaling during tracheal development in *Drosophila* (122,123). Within *C. elegans*,

mutations of the hst-6 gene resulted in viable and fertile offspring with neuron-specific defects in axon and cellular guidance (124,125). The effects seen with the 6OST-defective *C. elegans* have been linked to two important signaling systems, Slit/Robo and KAL-1 pathways, for nervous system development (124,125).

To date, there is no known crystal structure for 6OST however a structural neighbor search using mGenThreader has predicted that 6OST is most likely more structurally similar to NST and the 3OSTs instead of 2OST. The overall fold of the protein is most likely similar to NST and 3OST, however, there are some striking differences. The 6OST enzyme has ten evolutionary conserved cysteine residues while the NST and 3OSTs only possess two conserved cysteine residues whereas 2OST possesses four. Six of the cysteine residues within 6OST are located near the 5'-phosphate and 3'-phosphate groups of the proposed PAPS binding site where the other sulfotransferases have none in that site. The additional four cysteine residues are located after the 3'-phosphate of the PAPS binding site as is the case with 2OST. These four cysteine residues could potentially be important for the stability of 6OST as mutation of these same cysteine residues in 2OST destabilizes the protein (126). A combination of structural and extensive mutational analysis will guide our understanding of the substrate recognition mechanism of 6OST as well as contribute to our enzymatic approach to synthesizing structurally defined HS.

Glucosaminyl 3-O-Sulfotransferase (3OST)

The rarest modification to heparan sulfate is 3-*O*-sulfation of the glucosamine residue catalyzed by 3-*O*-sulfotransferase (Figure 12). The 3OST catalyzes transfer of the sulfo group from PAPS to the 3-*O*-position of either an unsubstituted or N-sulfated glucosamine residue within heparan sulfate (127). Studies have indicated that 3-*O*-sulfation accounts for

only 0.5% of the total sulfate within a heparan sulfate chain (128,129). Due to its infrequency, 3-*O*-sulfation of HS is of special interest since it has been identified to play a critical role in several biological processes including anticoagulation, herpes simplex virus (HSV) infection, circadian rhythms and various cancers of the breast, colon, lung and pancreas (127,130-132).

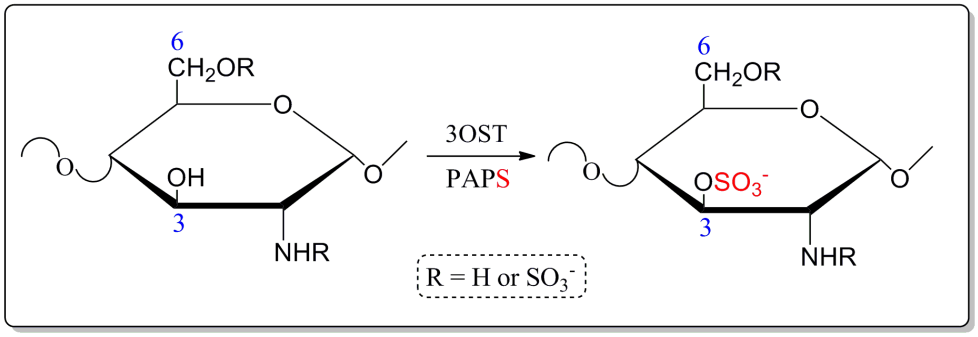


Figure 12. Reaction catalyzed by heparan sulfate 3OST. 3OST can transfer a sulfo group to either a $GlcNH_2$ or GlcNS residue to generate $GlcNH_23S$ and GlcNS3S respectively. The different 3OST isoforms are highly dependent on the non-reducing end residue adjacent to glucosamine.

Seven isoforms of 3OST have been identified in humans and mice and successfully cloned to date, consisting of 3OST-1, -2, -3_A, -3_B, -4, -5, and -6 (133-136). Based on sequence homology within the sulfotransferase domain, the isoforms can be divided into two subgroups, the 3OST-1-like and 3OST-3-like subgroups (137). The 3OST-1-like subgroup consists of 3OST-1 and 3OST-5, with approximately 75% sequence identity within the sulfotransferase domain (137,138). The 3OST-3-like subgroup consists of the other 3OST isoforms of 3OST-2, 3OST-3_A, 3OST-3_B, 3OST-4, and 3OST-6, possessing greater than 80% sequence identity within the sulfotransferase domain (137,138). Eight 3OST isoforms have been identified in zebrafish with members of the zebrafish family showing 63% identity with the corresponding isoform with the exception of the 3OST-5 isoform with 53%. The novel

zebrafish 3OST-7 gene was determined to be a member of the 3OST-1 like subgroup with 66% sequence identity to human 3OST-1 (137). A single 3OST gene, hst-3, has been identified in *C. elegans* (110) and two within *Drosophila*, 3OST-A and 3OST-B (138). The *Drosophila* 3OST-A is more homologous the 3OST-1 like subgroup whereas 3OST-B demonstrates higher identity with the 3OST-3 like group, suggesting that 3OST diverged into two branches with different specific functions during early evolution (138).

Despite the extensive characterization of the significance of 3-*O*-sulfation in the interaction of HS and antithrombin within anticoagulation, the role of 3OST in animal development has not been thoroughly investigated. The only available knockout model is that of 3OST-1 within mice, demonstrating normal hemostasis with abnormal phenotypes including genetic background-dependent lethality and intrauterine growth retardation (139,140). The interaction among HS and AT was unaffected in the knockout mice (139,140). The level of anticoagulantly active HS and factor Xa activity were reduced but not completely eliminated in the $3OST^{-/-}$ mice, suggesting that the other 3OST isoforms may compensate for the loss of 3OST-1 (139,140). These phenotypes are significantly distinct from AT knockout mice, which resulted in embryonic lethality due to thrombosis in both the liver and myocardium (141,142). The differences in the 3OST-1 and AT knockout mouse studies indicate that 3-*O*-sulfation may play a role in other biological processes and therefore more work must be carried out to explore this idea. A disruption of 3OST-B in *Drosophila* via RNA interference revealed a role for 3-*O*-sulfation in Notch signaling by markedly reducing the level of the Notch protein (138).

In contrast to the 6OSTs, the 3OST isoforms have very distinct substrate specificities, recognizing unique disaccharide structures within HS and generating HS with diverse

biological functions (Figure 13). The substrate requirements of 3OST-1, -3, and -5 have been extensively studied. The 3OST-1 enzyme transfers a sulfate to an *N*-sulfated glucosamine residue linked to a glucuronic acid or iduronic acid at the non-reducing end to generate GlcUA-GlcNS3S±6S and IdoUA-GlcNS3S±6S respectively. On the other hand, the 3OST-3 enzyme can sulfate two substrates, IdoUA2S-GlcNH$_2$ and IdoUA2S-GlcNH$_2$6S, which were found within the highly sulfated NS domains of heparan sulfate (135). Studies have determined that substrates modified by members of the 3OST-3-like family, including 3OST-2, 3OST-3, 3OST-4, and 3OST-6 do not possess anticoagulant activity as the HS modified by 3OST-1 but instead act as entry receptors for herpes simplex virus (HSV) 1 entry (134-136,144,145). The 3OST-3 enzyme can also recognize disaccharides carrying *N*-sulfated glucosamine linked to an IdoUA2S residue at the non-reducing end (143). The other members of the 3OST-3-like subgroup, 3OST-2, 3OST-4, and 3OST-6, were determined to recognize only disaccharides containing *N*-sulfated glucosamine linked to a 2-*O*-sulfated IdoUA at the non-reducing end (IdoUA2S-GlcNS±6S) (136,143,144,146) .

Interestingly, the 3OST-5 is very promiscuous in that it can recognize and sulfate the substrates for both 3OST-1 and 3OST-3, which includes recognition of glucosamine residues attached to a glucuronic acid, iduronic acid, or even a 2-*O*-sulfated iduronic acid at the non-reducing end (147). As a result of its broad specificity, HS modified by 3OST-5 can possess either anticoagulant activity or serve as an entry receptor for HSV-1 infection (145,147). Disaccharide analysis of both 3OST-1 and 3OST-5 modified HS isolated from CHO cells revealed the presence of a rare disaccharide of IdoUA-GlcNS3S±6S whose function is currently unknown (147-149).

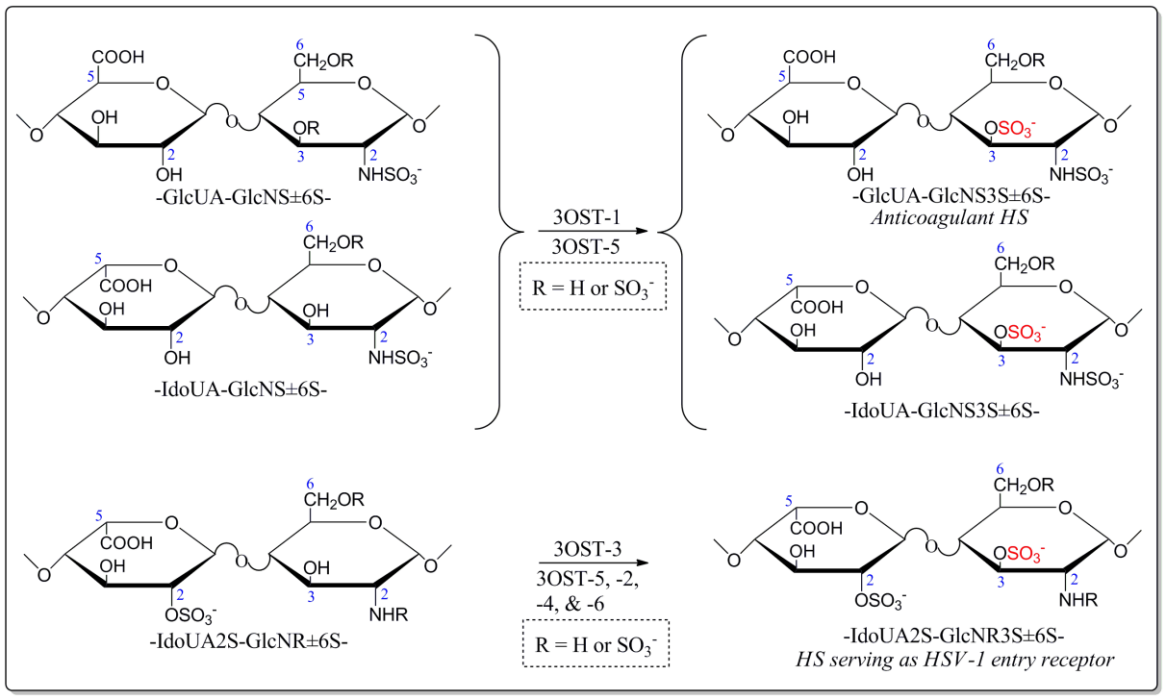


Figure 13. Substrate specificity of the 3OST isoforms. 3OST-1 sulfates the 3-OH position of a GlcNS residue linked to either GlcUA or IdoUA at the non-reducing end. 3OST-3 sulfates the 3-OH position of a GlcNS or $GlcNH_2$ linked to an IdoUA2S residue at the non-reducing end. 3OST-2, -4, and -6 can modify a GlcNS residue with an IdoUA2S residue at the non-reducing end that serves as an entry receptor for herpes simplex virus 1. 3OST-5 is promiscuous and can sulfate the substrates of both 3OST-1 and 3OST-3. The substrate modified by 3OST-1 possesses anticoagulant activity while the substrate modified by 3OST-3 is an entry receptor for herpes simplex virus 1. The 3OST-5 modified HS can demonstrate either anticoagulant activity or serve as an HSV entry receptor. The 3-*O*-sulfo group is colored red.

The crystal structures of the sulfotransferase domains of 3OST-1, 3OST-3, and 3OST-5 have been elucidated (63,150,151), providing a plethora of information regarding the catalytic and substrate recognition mechanisms of 3OST. The sulfotransferase domain of the 3OST isoforms were determined to be structurally similar to NST-1 with the exception of distinguishing features within the HS binding cleft. The 3OSTs have an overall folded structure with a large open HS binding cleft. Like NST-1, the 3OSTs have a central α/β motif. The PAPS binding site within the 3OSTs is nearly identical among the 3OSTs and NST-1, relying on a conserved glutamate as the catalytic base. The striking difference

between NST-1 and 3OSTs is that the 3OSTs have an increased positive charge within the substrate binding cleft, suggesting recognition of a highly negatively charged HS substrate (150). This finding may explain some of the differences in substrate specificity for these two biosynthetic enzymes.

The knowledge regarding the substrate recognition mechanisms for the 3OSTs was further advanced by the revelation of a ternary complex of 3OST-3 with PAP and a known 3OST-3 HS substrate, ΔUA2S-GlcNS6S-IdoUA2S-GlcNS6S (63,143) (Figure 14). It should be noted that the $\Delta^{4,5}$ unsaturated 2-*O*-sulfo uronic acid (ΔUA2S) is not present in the HS polysaccharide however the skew boat conformations of IdoUA2S and ΔUA2S present similar positioning of functional groups. Therefore, the ΔUA2S is a sufficient mimic of IdoUA2S for recognition by 3OST-3 (63). The tetrasaccharide binding site of 3OST-3 demonstrated extensive interactions with the non-reducing end UA2S unit as well as the identified the 3-OH position of non-reducing end glucosamine residue as the site of 3-*O*-sulfation (63). The amino acid residues (Lys-161, Arg-166, Lys-215, Gln-215, Lys-368, and Arg-370) within 3OST-3 involved in HS binding are conserved across the isoforms, suggesting an important role for these residues in dictating substrate binding (2,63). Mutation of these potential substrate binding residues within 3OST-3 resulted in a drastic loss of enzymatic activity, confirming their importance in substrate recognition (63). Interestingly, two 3OST-3 mutants, Q255A and K368A, nearly abolished sulfotransferase activity however mutation of the corresponding residues in 3OST-1 (Gln-166 and Lys-274) maintained 34% and 17% activity respectively (63). These results suggest that Gln-255 and Lys-368 are important for the substrate specificity of 3OST-3 since these residues were found to interact with UA2S and IdoUA2S residues within the bound tetrasaccharide. These

findings can also be supported by the fact that 3OST-1 recognizes a GlcUA instead of an IdoUA2S residue as a substrate. The information gained from the crystal structure of 3OST-3 with a bound tetrasaccharide provided insight into the substrate specificity of 3OST however crystal structures of the other isoforms with appropriate oligosaccharides are necessary for a clear picture of the substrate recognition mechanism of the 3OSTs.

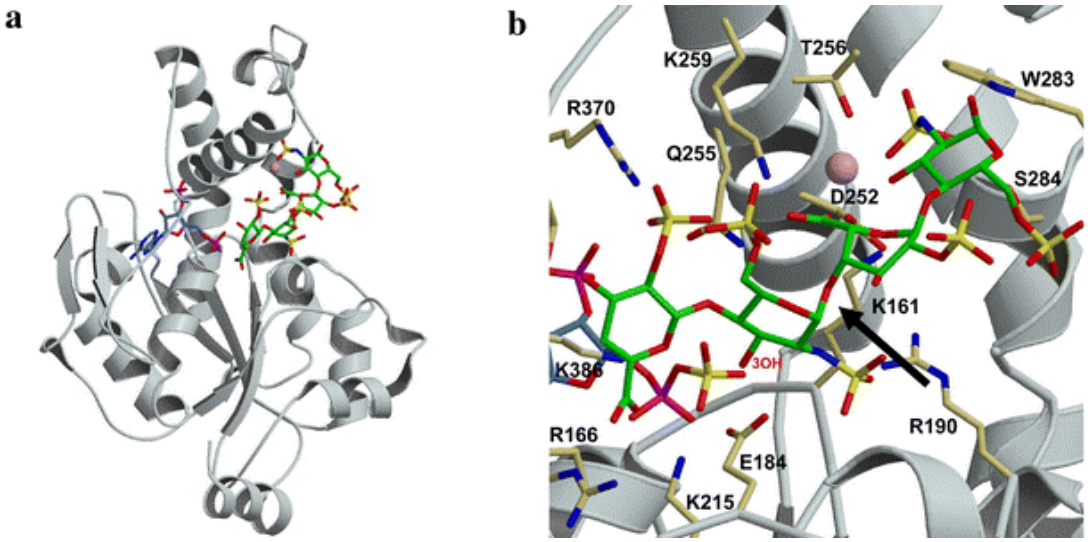


Figure 14. Crystal structure of ternary complex 3-OST-3/PAP/tetrasaccharide. (A) Crystal structure of 3OST-3A in complex with PAP (blue) and HS tetrasaccharide (green). **(B)** Superposition of PAPS onto PAP in the active site of 3OST-3A. The relative orientation of the acceptor 3'-OH position of the HS substrate is displayed. The side chains that are involved in substrate binding are labeled. A sodium ion involved in binding is labeled pink. The arrow indicates a kink in the polysaccharide. Structure taken from (152).

With an initial attempt to engineer the HS biosynthetic enzymes, the crystal structure of 3OST-5 was determined and successfully demonstrated the feasibility for tailoring the substrate specificity of the 3OST isoforms (Figure 15A) (151). As mentioned previously, 3OST-5 has been determined to possess both 3OST-1 (anticoagulant) and 3OST-3 (HSV-1 viral entry) activities therefore a structurally-guided mutagenesis study was completed to selectively remove the 3OST-1-like and 3OST-3 like activities of 3OST-5. As seen in Figure

15B&C, a gate was found to be formed by specific residues along the non-reducing end of the binding clefts of 3OST-1 (Glu-88 and His-271) and 3OST-3 (Gly-182 and Gly-365) and speculated to possibly confer the specificity of the enzymes. The gate residues form a narrow cleft for 3OST-1 whereas a much wider cleft is present for 3OST-3. 3OST-5 was determined to be more structurally similar to 3OST-3 with the gate residues (Ser-120 and Ala-306) forming a wider cleft (Figure 15D). Mutation of the corresponding gate residues in 3OST-5 to the 3OST-1 gate residues altered the specificity of 3OST-5 and vice versa. For example, mutation of the 3OST-1 gate residues to the corresponding 3OST-5 gate residues increased recognition for the IdoUA2S-GlcNS containing polysaccharide, a HS structure normally recognized by 3OST-3 and 3OST-5 instead of 3OST-1. Additional support was provided when mutation of the 3OST-5 gate residues to the 3OST-1 gate residues reduced the recognition for the IdoUA2S-GlcNS containing substrate, a known 3OST-3 substrate, while increasing recognition of the known 3OST-1 substrate containing GlcUA-GlcNS. The results suggested that the gate residues among the 3OST isoforms along with catalytic site are responsible for recognizing the appropriate polysaccharide substrate, specifically the presence of the IdoUA2S unit, a distance of three saccharide residues away from the site of sulfation (151). The information gained from the 3OST-5 structural analysis demonstrated the potential for engineering the sulfotransferases to generate HS with desired biological functions.

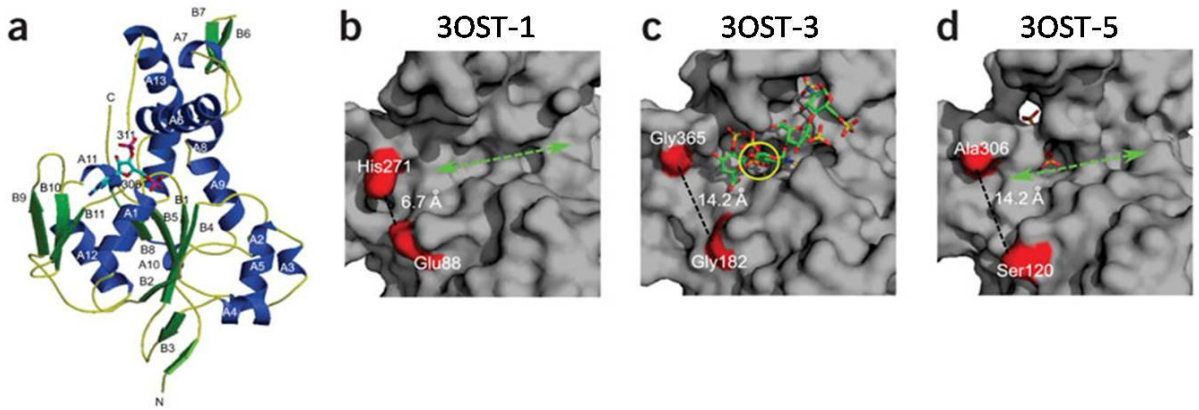


Figure 15. Structural analysis of the gate residues within the 3OST isoforms. (A) Ribbon diagram of the crystal structure of human 3OST-5 with PAP bound (cyan). **(B-D)** Identification of the substrate binding gate residue with the 3OST isoforms. The position of the substrate binding cleft is shown as a dashed green arrow. The measured distances between the gate residues are labeled black. The key amino acid residues forming a gate at the non-reducing end of the binding cleft of mouse 3OST-1 **(B)**, human 3OST-3 **(C)**, and human 3OST-5 **(D)** are labeled and the surface colored in red. The bound tetrasaccharide substrate, ΔUA2S-GlcNS6S-IdoUA2S-GlcNS6S, is depicted in ball and stick and the glucosamine residue to be sulfated is circled yellow. Structure taken from (151).

Current Methodologies for Structural Analysis of Heparan Sulfate

There is accumulating evidence explaining the role of heparan sulfate polysaccharides in regulating biological functions in both normal and disease states. Probing the structure-function relationship of heparan sulfate polysaccharides is essential for understanding their biological roles however the structural heterogeneity, complexity and anionic nature within the chains make sequencing of the polysaccharides difficult. Another barrier to sequencing is that the non-template based synthesis of the polysaccharides prevents amplification as is possible with DNA and proteins (4,153).

The most common analytical technique for determining the structural characteristics, which include acetylation, sulfation, and epimerization state, of HS polymers includes cleavage of the polysaccharide to disaccharides followed by reverse phase ion-pairing high performance liquid chromatography (RPIP-HPLC) or capillary electrophoresis to characterize the disaccharide species present in a cleaved polysaccharide sample (4). RPIP-

HPLC is a very effective technique because simply altering the chromatography conditions allows for determination of disaccharides with different sulfation states ranging from non-sulfated to tetrasulfated as well as epimerization states. The major limitation to this analytical technique is that only information regarding disaccharide fragments is provided and no decipherable information is available to identify how these fragments are pieced together to form the polysaccharide. Additional analytical techniques including mass spectrometry can be employed to determine the number of sulfo groups as well as the length of the oligosaccharide chain (146,153).

There are currently two cleavage methods for completing disaccharide analysis for HS compounds in our laboratory: (1) chemical degradation with nitrous acid and (2) enzymatic degradation with heparin lyases. Both depolymerization techniques cleave the glycosidic linkage between a non-reducing glucosamine residue and a reducing end uronic acid residue via an elimination reaction (15).

Chemical Degradation Using Nitrous Acid

Depolymerization of HS polysaccharides using nitrous acid offers an effective method for determining the sulfation types within a HS chain. The use of nitrous acid serves to cleave HS between a non-reducing end glucosamine residue and reducing end uronic acid to produce a disaccharide containing a non-reducing end uronic acid and a reducing terminal 2,5-anhydromannose, which can be reduced to anhydromannitol (AnMan) using sodium borohydride (Figure 16) (15). The cleavage reaction with the glucosamine residue is highly dependent on pH. The glycosidic bonds of *N*-sulfo glucosamine residues react with nitrous acid at pH 1.5 whereas the bonds of *N*-unsubstituted glucosamine residues are cleaved when treated with nitrous acid at pH 4.5-5.5 (4,15). Unfortunately, the *N*-acetyl glucosamine

residues do not react with nitrous acid at either pH and must first be converted to *N*-unsubstituted glucosamine ($GlcNH_2$) using hydrazine, allowing for subsequent cleavage with nitrous acid at pH 4.5. The advantage of this method is that the configuration of the uronic acid is unaltered, allowing one to distinguish between IdoUA and GlcUA. The major disadvantage is the need for the use of a radioactive tag or fluorescent label must be incorporated prior to disaccharide analysis since the disaccharides do not possess intrinsic UV absorbance (4).

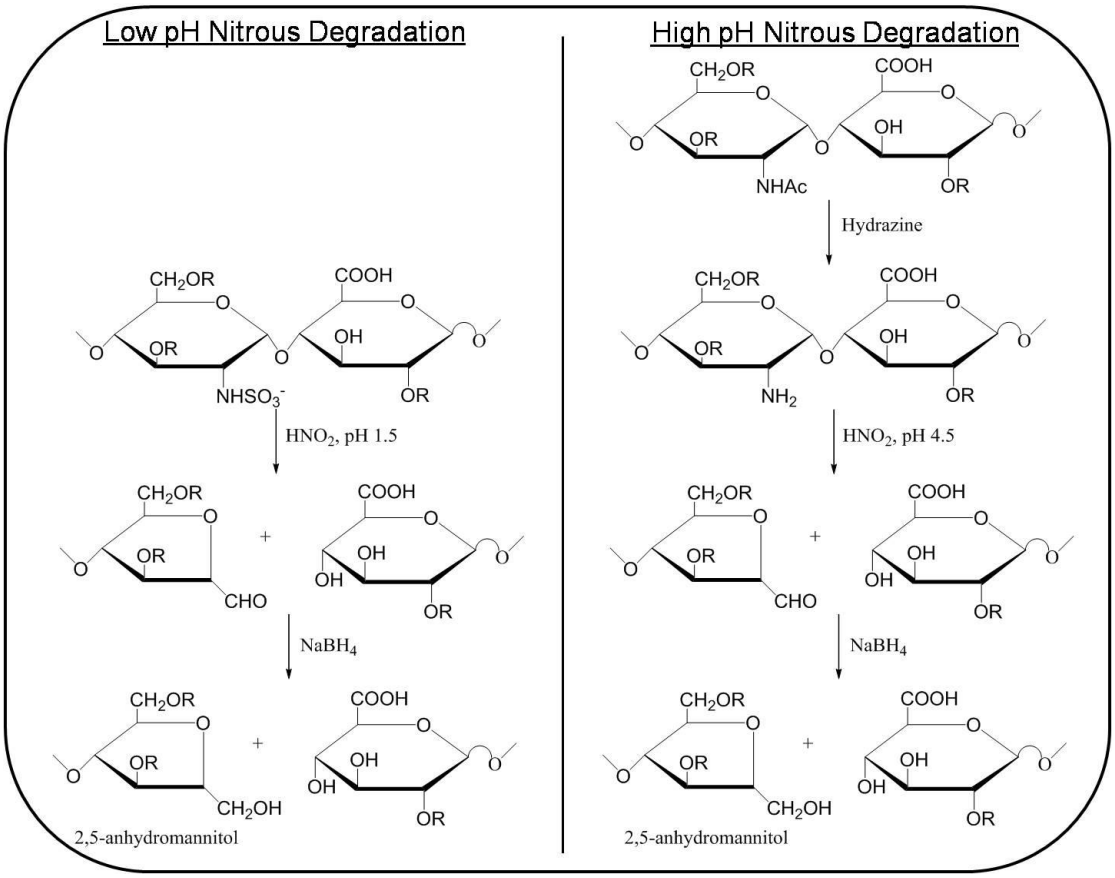


Figure 16. Nitrous acid cleavage of heparan sulfate. At pH 1.5, nitrous acid cleaves between aGlcNS residue and uronic acid. Nitrous acid cleaves between $GlcNH_2$ residues and uronic acid at pH 4.5. The GlcNAc residue must first be hydrolyzed to $GlcNH_2$ using hydrazine followed by nitrous acid cleavage at pH 4.5. Adapted from (4).

Enzymatic Degradation Using Heparin Lyases

Enzymatic degradation has been utilized extensively for the characterization of HS polysaccharides. This degradation technique employs heparin lyases, originally purified from *Flavobacterium heparinum*. There are three heparin lyases, I-III, which have been determined to possess slightly different substrate requirements for cleavage of the glycosidic bond between glucosamine and uronic acid via a β-elimination reaction (Figure 17). Heparin lyase I cleaves the glycosidic bond between an *N*-sulfo glucosamine residue and 2-*O*-sulfated iduronic acid. Heparin lyase III recognizes the linkage between either *N*-acetylated or *N*-sulfated glucosamine and unsulfated glucuronic acid. On the other hand, heparin lyase II is much more promiscuous in that it recognizes the substrates for both lyases I and II, cleaving between GlcNS/GlcNAc and GlcUA/IdoUA (15,153). For each lyase, the catalyzed reaction generates a $\Delta^{4,5}$-unsaturated uronic acid at the non-reducing end of the product that possesses an intrinsic UV absorbance at 232 nm, allowing for easy monitoring with RPIP-HPLC techniques. The major limitation to the enzymatic degradation is the loss of the configuration of the C_5 position of uronic acid residue upon formation of the $\Delta^{4,5}$-double bond. This prevents discrimination between the IdoUA and GlcUA residues within the product. Noteworthy, exhaustive digestion of a mixture of the lyases does not completely cleave the polysaccharide or oligosaccharide to disaccharides. Therefore, a significant amount of material remains as oligosaccharides which cannot be detected by the separation techniques, resulting in an underestimation of amount of sample present (4).

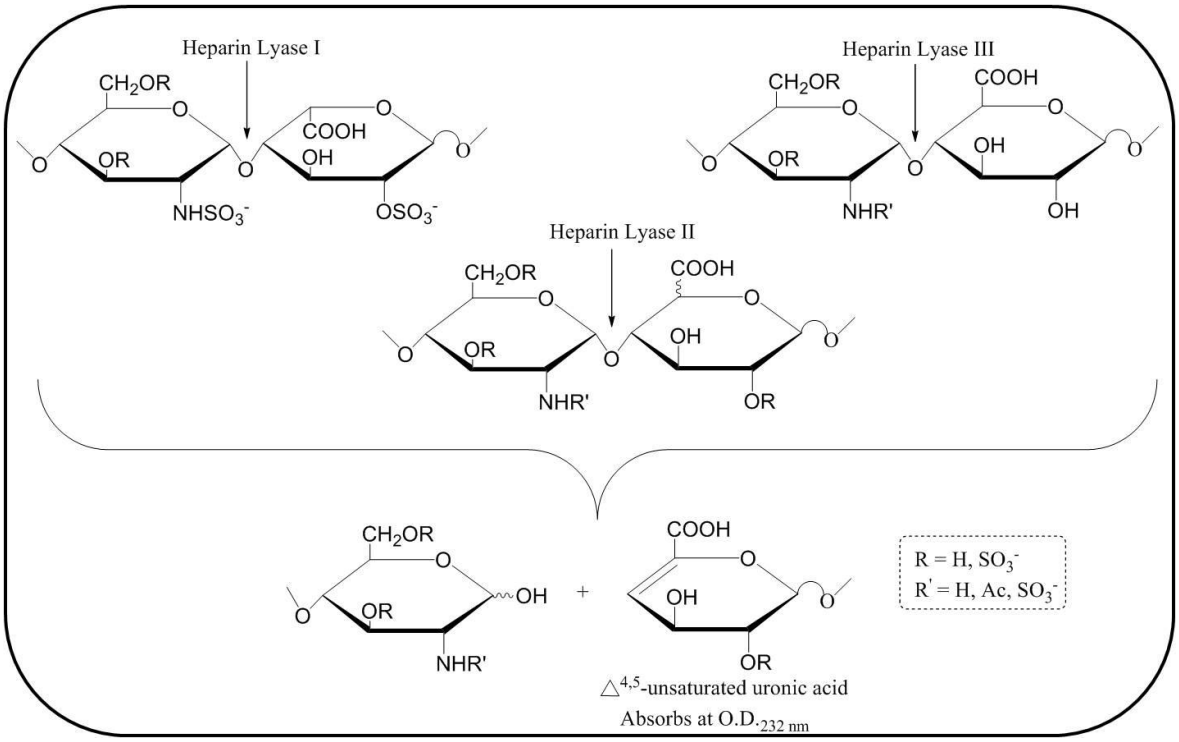


Figure 17. Substrate specificity of the heparin lyases. Heparin lyase I cleaves between GlcNS and IdoUA2S whereas heparin lyase III cleaves between GlcNAc/GlcNS and unsulfated glucuronic acid. Lyase II can cleave the substrates of both I and III. Adapted from (4)

Physiological and Pathophysiological Functions of Heparan Sulfate

The modifications implemented by the HS sulfotransferases and epimerase are directly involved in contributing to the structural diversity of heparan sulfate. The structural diversity of HS allows for interaction with a variety of biologically relevant proteins, suggesting an important regulatory role for HS in different physiological and pathophysiological functions. The structure-function relationship of HS with such proteins as antithrombin, fibroblast growth factors, and glycoprotein D has been extensively characterized.

Anticoagulation

The blood coagulation pathway is essential for both hemostasis and thrombosis, consisting of a proteolytic cascade resulting in formation of fibrin thrombi, or a blood clot (Figure 18) (152). The cascade consists of several serine proteases existing as inactive zymogens that are cleaved and activated by an upstream protease, serving to catalyze the activation of the subsequent reaction in the cascade. The blood coagulation cascade can be divided into two separate pathways, intrinsic and extrinsic. The intrinsic pathway is triggered by contact activation of plasma factor XII when exposed to a damaged surface. The primary extrinsic pathway is initiated at a site of injury in response to the release of tissue factor. Although distinctly initiated, the two pathways do converge to a single pathway known as the factor X activation pathway. Once factor Xa serves to activate thrombin, its primary role is to convert fibrinogen to fibrin. Fibrin then forms cross-links to generate a blood clot (154).

Antithrombin-III (AT) is a key regulatory protein in the coagulation cascade that serves to maintain proper blood flow and prevent thrombosis (155). The role of AT in embryogenesis was established using an *in vivo* mice knockout model, which resulted in embryonic lethality and subcutaneous hemorrhage. The AT deficient mice exhibited excessive fibrin deposition in the myocardium and liver, suggesting an important regulatory role for AT (142).

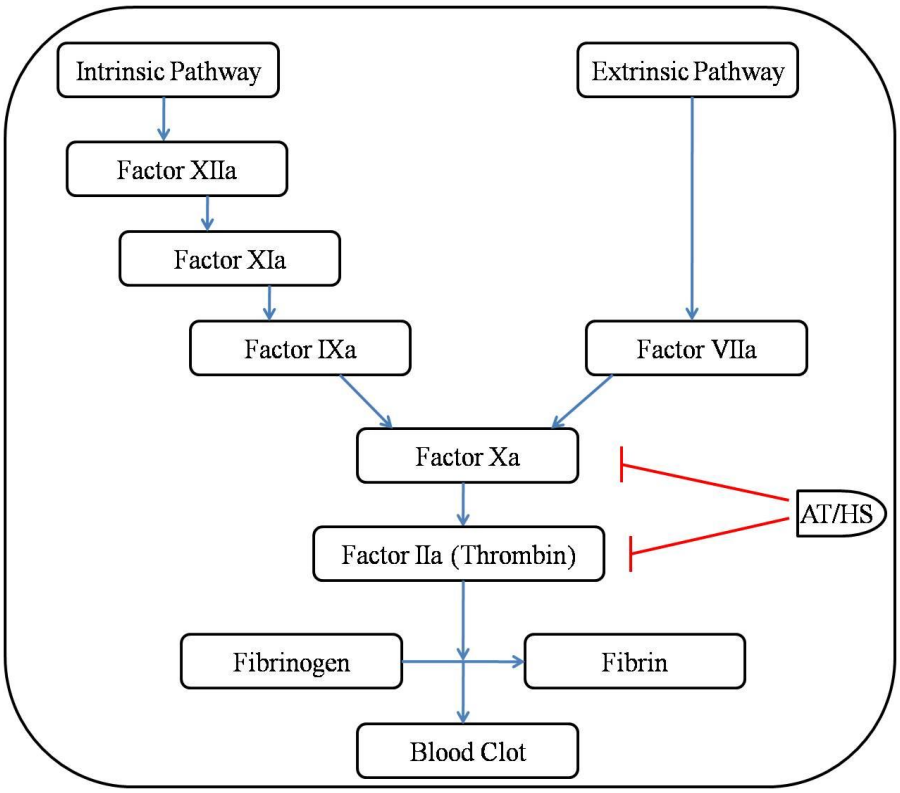


Figure 18. Blood coagulation cascade. The coagulation process includes a series of inactive proteases that are cleaved, activated, and then proceed to activate the subsequent zymogen. The cascade is initiated by two pathways, intrinsic and extrinsic pathways. The two pathways converge into one pathway for factor X activation ultimately leading to the formation of a blood clot. The "a" denotes the activated from of each protease. Coagulation can be counteracted by the inhibitory action of an AT/HS complex (indicated by the *red* lines) against both factor Xa and thrombin.

The coagulant activity of the blood coagulation pathway can be counteracted by the binding of HS/HEPARIN to AT. Although AT has an innate ability to inhibit thrombin and factor Xa, the inhibitory activity is significantly enhanced 9,000 fold for thrombin and 17,000 fold for factor Xa by the binding of heparin (156). Binding of HS/HEPARIN to AT results in a conformational change of AT to an active form that can inhibit blood coagulation factors Xa and IIa (thrombin) (12,15,152). Heparin, the product of degranulated mast cells, has a much stronger anticoagulant activity than endothelial HS with HS possessing only 1-10% anticoagulant activity (152). Heparin can inhibit both factor Xa and thrombin for complete

anticoagulant activity, possibly explaining the discrepancy among HS and heparin activity (152,157,158).

Extensive structural and functional studies have identified a pentasaccharide domain within both HS and heparin that is crucial for AT binding (12,159) (Figure 19). Studies have revealed that 3-*O*-sulfation of the internal *N*-sulfo glucosamine residue is essential for AT binding, demonstrating a 20,000 fold decrease in activity when removed (160). 6-*O*-sulfation at the nonreducing end, which can be either GlcNAc or GlcNS) of the HS pentasaccharide has also been shown to be important with removal resulting in complete abolition of AT binding (161). It was originally determined that the IdoUA2S at the reducing end of the 3-*O*-sulfated glucosamine residue was essential for the binding of heparin to AT however recent studies have indicated that neither the IdoUA residue nor 2-*O*-sulfation of the IdoUA is necessary (162). It was also suggested that the IdoUA residue is necessary for AT binding when the oligosaccharide is smaller than a hexasaccharide but inessential once the oligosaccharide becomes larger in size (162). Therefore, elimination of these unnecessary features could aid in the synthesis of an anticoagulant agent with reduced side effects.

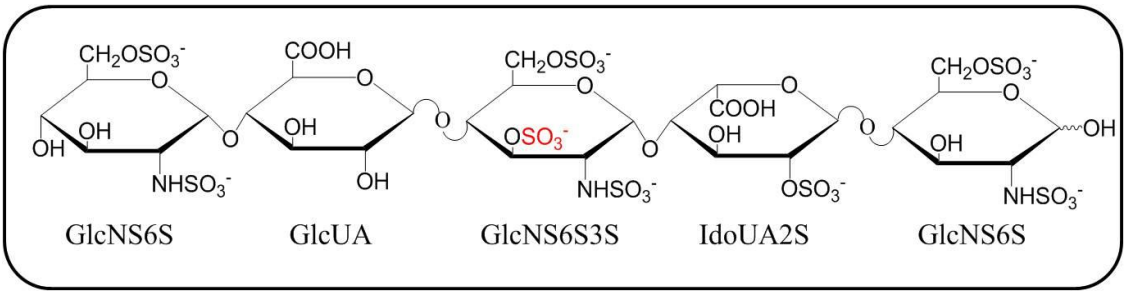


Figure 19. Antithrombin binding heparan sulfate pentasaccharide. The critical 3-*O*-sulfo group is colored *red* for emphasis. The abbreviated name is listed under each monosaccharide unit. The HS pentasaccharide binds antithrombin with a K_d of 15-20 nM.

The AT binding pentasaccharide was chemically synthesized and marketed as the antithrombotic drug Arixtra in 2001; this pentasaccharide only has the ability to inhibit factor Xa activity (163). A longer polysaccharide (approximately 14-20 saccharides) is necessary to inhibit the activities of both factors Xa and thrombin for complete anticoagulant activity (164). For thrombin inhibition, the longer polysaccharide must contain a negatively charged template to allow formation of a ternary complex with both AT and heparin (165). Chemical synthesis of the pentasaccharide takes more than 60 steps however Kuberan *et al.* were able to rely on the HS biosynthetic enzymes to synthesize the pentasaccharide in solely 6 steps with a much higher yield. The efficiency of enzymatic synthesis offers a promising approach to design structurally defined HS compounds with distinct biological functions.

Cell Proliferation

The fibroblast growth factor (FGF) family is involved in many important biological processing including cell proliferation, differentiation and migration during morphogenesis as well as regulating tissue repair, wound healing, and tumor angiogenesis among adult organisms (166). Currently, 22 FGFs have been identified to date and they all have been established as having affinity for both heparin and HS (167). These growth factors mediate their cellular responses through binding and activation of a family of receptor tyrosine kinases known as FGF-receptors (FGFR) (166). Heparan sulfate functions to enhance the affinity between FGF and FGFR as well as promote the dimerization and activation of FGFRs (168). Structural studies have focused on two members of the FGF family, namely FGF1 and FGF2. Crystal and co-crystal structures have been solved for FGFs bound to different HS oligosaccharides and FGFRs (168-170).

Structural analysis using surface plasmon resonance, NMR, and mass spectrometry has implicated a disaccharide sequence possessing several sulfate groups that are important for linking FGF and FGFR. The 2-*O*-sulfo group on IdoUA and *N*-sulfo group on glucosamine are necessary for both FGF1 and FGF2 binding and signaling. The 6-*O*-sulfo group on glucosamine is essential for binding and activation of FGF1 but not for FGF2 binding (122,171). It was also shown that saccharide length plays a crucial role in promoting dimerization (122,171-173). Maccarana *et al.* identified the minimal heparin structure responsible for binding to FGF-2 as a pentasaccharide with the sequence ΔUA-GlcNS-UA-GlcNS-IdoUA2S (Figure 20) (2,174). The FGF-1 binding site, on the other hand, requires at least a penta- to hexasaccharide and contains a critical IdoUA2S-GlcNS6S disaccharide unit (2,175). Chen *et al.* explored the effects of FGF-2 mediated cell proliferation by enzymatically modifying completely de-*O*-sulfated *N*-sulfated heparin with a combination of different HS biosynthetic enzymes to generate different sulfation patterns. The results of this study demonstrated that the compounds with both 2-*O*- and 6-*O*-sulfations were capable of binding FGF-2 and promote cell proliferation compared to heparin, suggesting that these modifications are essential for promoting FGF-2-mediated cell proliferation on a polysaccharide level (176).

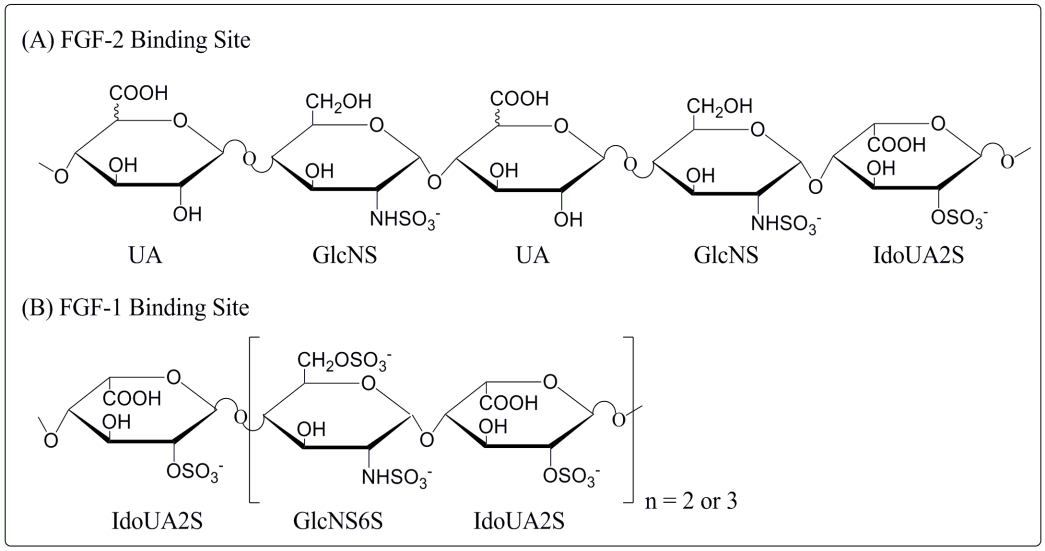


Figure 20. Heparan sulfate structures implicated in FGF-1 and FGF-2 binding. Adapted from (2)

Until recently, it was believed that a specific sulfation pattern on heparan sulfate was responsible for the interaction between HS and growth factors based on *in vitro* studies (177). However, recent *in vivo* studies using *Drosophila* Hs2st and Hs6st mutants revealed a compensatory increase in sulfation to maintain the overall negative charge of the HS. This suggests that charge density is more important than HS structures with precisely positioned sulfate groups for mediating FGF signaling (178). A detailed understanding of the mechanism of FGF binding and activation by HS will help target the suppression of tumor angiogenesis as an anti-cancer approach.

Inflammation

Heparan sulfate is a key player in the regulation of leukocyte transmigration during inflammation through interaction with various chemokines and cytokines. A hallmark of the inflammatory response includes leukocyte recruitment to the site of infection (179). The recruitment is initiated by the binding of leukocytes to different lectins (L-selectin and P-

selectin) expressed on the surface of the endothelium that allow for leukocytes to roll along the surface. Studies have indicated that both heparin and HS interact with P- and L-selectins, interfering with selectin-mediated leukocyte rolling (180,181). The removal of HS from the aortic endothelial cell surface resulted in inhibition of L-selectin mediated binding of monocytes *in vitro* (181). Independent of anticoagulant activity (182), the use of heparin and heparin-like oligosaccharides demonstrated anti-inflammatory activity by blocking both P- and L-selectin binding (183,184).

Although HS has been shown to interact with both P- and L-selectins, it is not known whether it is the ligand under physiological conditions since these selectins can also interact with mucins. Studies have indicated that heparin and HS have a lower affinity for P-selectin than L-selectin, suggesting that L-selectin is a natural substrate (183). The L-selectin can interact with sulfated, sialylated, and fucosylated mucins on the cell surface, and these mucins are speculated to be the ligand during chronic inflammation (185,186). However, during the acute inflammatory response, studies have indicated that HS is the dominant L-selectin ligand (187). The inactivation of the NDST enzyme in endothelial cells resulted in partially sulfated HS that presented a significant reduction in the recruitment of neutrophils to inflamed tissues (187). That reduction was due in part to the lack of L-selectin binding to HS. It was also demonstrated that partially sulfated HS on the surface of leukocytes did not affect recruitment to inflamed tissues (187). Despite these contradictory results, it is still clear that HS plays an important role during the inflammatory response.

Endothelial HS also plays a role in the interaction of chemokines with their cognate receptors on the surface of neutrophils as well as chemokine transportation (i.e. transcytosis) across the endothelial cell barrier (188,189). Heparan sulfate interaction with chemokines

provides protection from proteolysis and stimulates chemokine oligomerization (179,190,191). This interaction allows for immobilization of chemokines on the endothelial surface, which must occur in order for leukocytes to interact with the endothelium and migrate. This theory was supported by a recent study showing immobilized chemokines induced extension of the lymphocyte functional-associated antigen 1 (LFA-1), an integrin that promotes stable cell adhesion (192). The role of heparan sulfate in promoting chemokine transcytosis has also been established. Macrophages and mast cells release chemokines during an inflammatory response that must be transported to the endothelial cell barrier to be displayed on the endothelial cell surface. These transported chemokines are normally bound to heparan sulfate, suggesting that HS may assist with this process. An *in vivo* study using HS deficient mice demonstrated that the efficiency of mediating chemokine transcytosis is significantly reduced, suggesting that HS is indeed involved in transcytosis however the mechanism is still poorly understood (179,187).

Since heparan sulfate plays a role in many stages of the inflammatory process, it seems potentially favorable to generate inhibitors of heparan sulfate function as an anti-inflammatory therapy (179). It has long been established that heparin possesses anti-inflammatory activity (193-196). However, its structural diversity has been implicated in a variety of undesired side effects. Therefore, it could be possible to develop heparin or HS mimetics that target leukocyte transmigration as anti-inflammatory therapeutics.

Viral Infection

The initial step for propagation of viral infection includes binding of a viral particle to target cells. Heparan sulfate has been shown to serve as a docking site for viruses through interaction of viral envelope proteins and facilitates viral internalization (4,130,197). Since

heparan sulfate is present on the surface of virtually all mammalian cells, it is not surprising that a multitude of human pathogenic viruses rely on interactions with HS including human immunodeficiency virus (HIV) (198), hepatitis C (199), human papillomavirus (HPV) (200), and herpes simplex virus-1 (HSV-1) (4,197).

The role of HS in mediating HSV-1 infection, hallmarked by facial mucocutaneous lesions and keratitis, has been extensively studied (201,202). HSV-1 infection occurs in two stages: attachment to cells and entry into cells (Figure 21) (202,203). The initial binding of HSV-1 to target cells occurs through interaction of viral envelope glycoprotein C (gC) and sometimes glycoprotein B (gB) with cell surface HS, labeled as the binding receptor (204,205) (206). The removal of cell surface HS prevented HSV-1 binding and rendered the cells partially resistant to HSV-1 infection, implicating its role in assisting HSV-1 attachment to target cells (202). Structural studies of the HS involved in binding to gC have indicated that at least 10-12 saccharide units containing both IdoUA2S units and 6-*O*-sulfated glucosamine (GlcNAc6S/GlcNS6S) are essential for interaction (207).

Once the HSV-1 is attached to the target cell, the viral envelope must fuse with the host membrane to allow for viral entry (4). The fusion between the viral envelope and cell membrane is triggered by multiple receptor-ligand mediated interactions at the cell surface (201). *In vitro* studies have indicated that HSV-1 utilizes three specific host cell surface receptors for viral entry: (1) herpesvirus entry mediator (HVEM), a membrane of the tumor necrosis factor family, (2) nectin 2, and (3) nectin 1 (4,208,209). The nectin receptors are members of the immunoglobulin family. Each of these three receptors relies on interaction with the viral envelope glycoprotein D (gD), possessing binding in the low micromolar

range, however this single interaction is not enough to trigger fusion (210-213). There are other key viral glycoproteins, gB, gH, and gL, that are important for the fusion process (210).

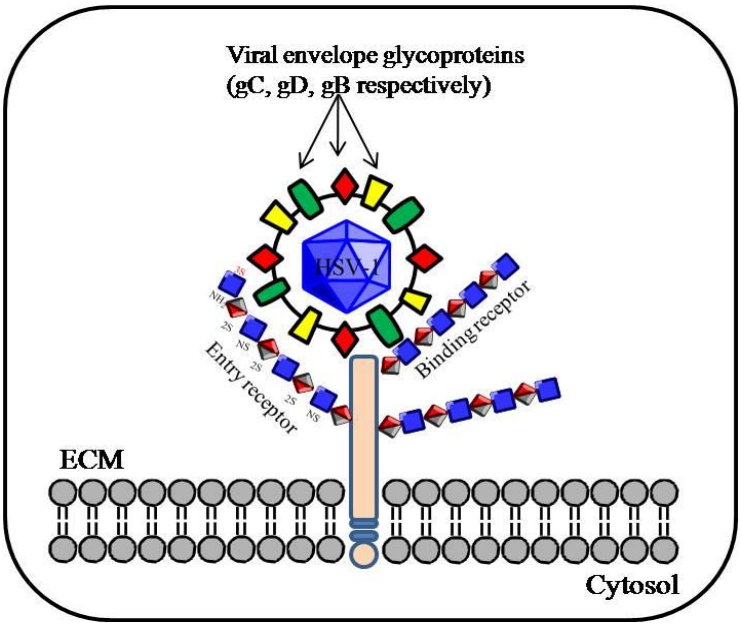


Figure 21. Interactions between a HSV-1 viral particle and the cell surface HS binding and entry receptors. Cell surface heparan sulfate, containing IdoUA2S and 6-*O*-sulfated glucosamine, are required for binding of the HSV-1 particle to the target cell through interaction with gC and sometimes gB. A specific 3-*O*-sulfated heparan sulfate serves as an entry receptor for HSV-1 through interaction with gD on the viral envelope**.**

Interestingly, Shukla *et al.* discovered a specific 3-*O*-sulfated HS acts as an entry receptor for HSV-1 infection through interaction with viral gD, predominately containing disaccharides of IdoUA2S-GlcNH$_2$3S±6S (130). This disaccharide is generated by 3OST-3 modified heparan sulfate but not those modified by 3OST-1. The implication of the 3-*O*-sulfated HS in HSV-1 entry was confirmed by the ability of HSV-1 resistant CHO cells to confer susceptibility to infection when 3OST-3 was introduced (214). In addition to 3OST-3, studies have indicated that several 3OST isoforms are responsible for generating the HS implicated in gD binding including 3OST-2, -4, -5, and -6 (63,134,136,143-145). The HS involved in binding and entry were determined to be different as unmodified HS isolated

from CHO cells is completely functional for promoting HSV-1 binding but not entry, suggesting that modification by specific 3OSTs is specific for generating HS involved in gD binding for viral entry (197). Characterization of a gD binding octasaccharide identified from a 3OST-3 modified oligosaccharide library was determined to possess the structure ΔUA-GlcNS-IdoUA2S-GlcNAc-UA2S-GlcNS-IdoUA2S-GlcNH$_2$6S3S (

Figure 22) (203). This oligosaccharide contains the previously identified signature disaccharide IdoUA2S-GlcNH$_2$6S3S necessary for viral gD binding. Identification of this gD binding octasaccharide enhanced our understanding of the role of HS in assisting HSV-1 infection.

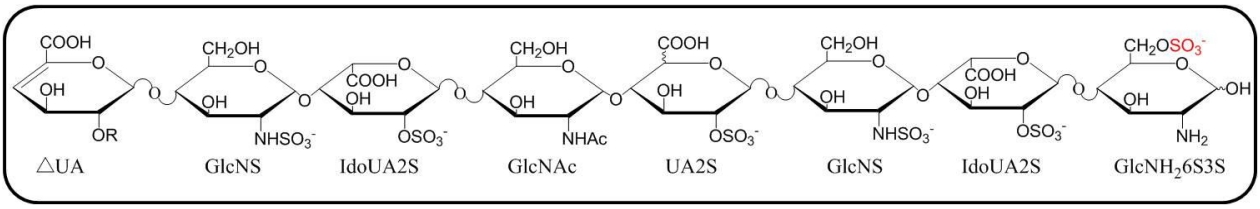


Figure 22. Structure of the gD binding heparan sulfate octasaccharide. The octasaccharide was determined to bind gD with a K_d of 18μM. The abbreviated name is labeled under each monosaccharide unit with the critical 3-O-sulfo group for gD binding colored red.

Cell surface heparan sulfate plays a role in several stages of infection for multiple viruses and therefore is an attractive therapeutic target for treating these infections. Targeting HS as an anti-viral agent could prevent the attachment and entry of viruses, reducing the risk of infection. After several attempts with sulfated polymers (215,216), Copeland *et al.* identified a HS mimetic of 3-*O*-sulfated heparin octasaccharide, ΔUA2S-GlcNS-IdoUA2S-GlcNS6S-IdoUA2S-GlcNS6S3S-IdoUA2S-GlcNS6S, that was capable of inhibiting HSV-1 infection with an IC_{50} of 40μM (217). This heparin octasaccharide actually

mimicked the gD binding site of 3-*O*-sulfated HS to completely block HSV-1 viral entry (217). It was previously established that heparin prevents HSV-1 infection, which is further confirmed by this study (197). Identification of the inhibitory heparin octasaccharide supports the potential for targeting 3-*O*-sulfated HS mimetics as an effective anti-viral therapy.

Tumor Progression

Heparan sulfate proteoglycans are present on the surface of tumor cells as well as cells involved in tumor survival, playing a role in both angiogenesis and tumor metastasis. HSPGs have the ability to interact with a variety of pro-angiogenic growth factors including vascular endothelial growth factor (VEGF), FGF-1/FGF-2 as well as anti-angiogenic factors such as endostatin (218). The binding site for endostatin was determined to be distinct from that of pro-angiogenic factors, and therefore endothelial cells may be modulating their HSPG profiles to become more or less sensitive to angiogenesis (218). The tumor cell surface HSPGs have the ability to contribute to tumor metastasis by recognizing and binding P-selectin, a ligand involved in cell adhesion to platelets or the endothelial lining. This binding to P-selectin allows for tumor cells to enter the bloodstream and invade other tissues (218,219).

Interestingly, heparin, a highly sulfated HS, has been shown to demonstrate anti-tumor effects when administered to cancer patients (220,221). Clinical studies have indicated that heparin may be beneficial for treatment of patients with small cell lung cancer (222). An *in vivo* mouse study indicated that administration of pharmacological doses of heparin inhibited tumor cell adhesion through inhibition of P-selectin binding (219,223). Thrombosis has been tightly associated with cancer as tumor cells are known to release

procoagulants such as tissue factors and cysteine proteases that serve to activate the blood coagulation pathway and form a fibrin shell around the tumor (224). The formation of this shell is speculated to protect the tumor from the immune response as well as cause resistance to chemotherapy, which can be interrupted by heparin administration (221).

Heparanase, an endo-β-D-glucuronidase, is known to cleave cell surface HS into small fragments, resulting in remodeling of the extracellular matrix (2,225). Heparanase cleavage has been attributed to tumor metastasis, and indeed heparanase is upregulated in numerous tumor cells including pancreatic tumors, esophageal carcinomas, and liver carcinomas (226). Therefore, heparanase is an excellent molecular target for the treatment of cancer (226). Studies have indicated that heparin and low molecular weight (LMW) heparin, a shorter version of heparin, is an inhibitor of heparanase (225,227,228). Despite its high inhibitory activity against heparanase, heparin is unsuitable for cancer treatment because of its potent anticoagulant activity. If the anti-cancer and anticoagulant activities could be separated, then heparin would offer a suitable treatment. The only issue is that patients may still require heparin for the treatment of thrombosis associated with cancer thus making heparin somewhat of a double-edged sword. The development of a structurally similar heparin mimetic possessing no anticoagulant activity could be a very effective therapeutic for cancer patients. *In vitro* studies analyzing different sulfated oligosaccharides against heparanase activity identified a promising heparin mimetic, phosphomannopentose sulfate (PI-88). PI-88 has an inhibitory activity against heparanase similar to heparin and clearly demonstrated promise in treating solid tumors during phase I clinical trials. The potential drug entered phase II clinical trials and demonstrated promise for treating post-resection liver patients. Despite the successful phase II clinical trials, the lack of a global partner caused

termination of multinational phase III clinical trials (229). Therefore, the search continues for an effective heparanase inhibitor for targeting angiogenesis and metastasis.

Based on the currently available research, heparin indeed plays a role in the prevention of tumor progression however it is unclear which mechanisms are responsible for this activity. There is certainly potential for the development of a heparin mimetic that can be utilized as an anti-cancer therapeutic.

Statement of Problem

Heparan sulfate (HS), a highly sulfated linear polysaccharide, is present on the surface and within the extracellular matrix of most animal cells. This polysaccharide regulates several functions of the blood vessel wall including blood coagulation, cell differentiation, and the inflammatory response. The wide range of biological functions makes HS an attractive target for the development of anti-cancer and anti-inflammatory agents. The functional diversity of HS is mediated by its structural heterogeneity. The backbone HS chain consists of repeating disaccharide units of glucuronic acid and *N*-acetyl glucosamine during biosynthesis. The HS backbone is susceptible to several modifications carried out by HS biosynthetic enzymes, generating uniquely sulfated saccharide sequences that determine the specific functions as seen by the anticoagulant activity of HS and heparin. The major route to preparing HS oligosaccharide structures is chemical synthesis however synthesis of HS oligosaccharides larger than a hexasaccharide is tedious and time-consuming therefore our laboratory plans to employ the use of heparan sulfate biosynthetic enzymes to synthesize homogeneous polysaccharides.

The biosynthesis of HS is a non-template driven process therefore understanding the mechanism of action and substrate specificity of the HS biosynthetic enzymes is critical for our approach to synthesizing biologically active HS compounds. The main goal of this dissertation was to characterize the structure-function relationship of two important HS *O*-sulfotransferases, specifically 2OST and 6OST. The combination of a crystal structure and elaborate mutational analysis should help provide some insight into the substrate recognition mechanism of these two enzymes.

Chapter II

Materials and Methods

Preparation of Competent Cells

A two µL aliquot of the competent cell stock was inoculated into 1 mL sterile LB medium containing appropriate antibiotics if necessary. For example, Origami B (DE3) competent cells have antibiotic resistance to both tetracycline and kanamycin. The cells were incubated at 37°C with shaking at 220 rpm overnight. A 1 mL aliquot of the saturated overnight culture was added to 100 mL sterile LB medium without any antibiotics. The cells were incubated at 37°C with shaking at 220 rpm until $O.D._{600}$ reached 0.5 after approximately 3 to 4 hours. The flask was then chilled on ice for 20 minutes and the cells were collected by centrifugation at 6000 rpm for 10 minutes at 4°C. The cells were then resuspended in 10 mL ice-cold filtered TSS (Transformation and Storage Solution) solution. The TSS solution was prepared fresh each time and contained 85% LB medium, 10% polyethylene glycol (wt/vol, mol wt. 8000), 5% dimethyl sulfoxide (DMSO), 50 mM $MgCl_2$, pH 6.5-6.6. The solution was filtered using a sterile 0.45 µM syringe filter (Millipore). Once resuspended, the cells were then aliquoted into microcentrifuge tubes, flash frozen, and stored at -80°C.

Chemical Transformation of Competent Cells

To transform competent cells, a 1-2 µL aliquot (100 ng) of plasmid DNA was added to 50 µL competent cells that were thawed on ice. The cells were incubated on ice for 30 minutes with occasional mixing. The cells were then heat shocked at 42°C for 40 seconds followed by incubation on ice for an additional 2 minutes. A 1 mL aliquot of sterile LB medium was added to the cells. The cells were then incubated at 37°C for 2-3 hours at 220 rpm. Following incubation, 100 µL of the transformation reaction was plated on LB plates with the appropriate antibiotics. The plate was incubated overnight at 37°C.

Expression of Heparan Sulfate Sulfotransferases in Origami B/Origami B^{chap} cell lines

Origami B (DE3) or Origami B^{chap} (Origami B cells expressing chaperonin proteins GroEL/GroES) cells were transformed using the protocol above. For Origami B (DE3) cells, a single colony was inoculated into 3 mL sterile LB medium 12.5 µg/mL tetracycline, 50 µg/mL carbenicillin (only if plasmid has resistance), and 15 µg/mL kanamycin and incubated at 37°C with shaking at 220 rpm for 3-4 hours. For Origami B^{chap} cells, a single colony was inoculated into 3 mL sterile LB medium 12.5 µg/mL tetracycline, 50 µg/mL carbenicillin (only if plasmid has resistance), 15 µg/mL kanamycin, and 40 µg/mL chloramphenicol incubated at 37°C with shaking at 220 rpm for 3-4 hours. Once the $O.D._{600}$ reached 0.6-0.8, the temperature was reduced to 22°C to cool down the culture. For the Origami B (DE3) cells, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.2 mM and the cells were allowed to shake overnight at 22°C. For Origami B^{chap} cells, 1 mg/mL L-arabinose was added first to induce expression of the chaperonin proteins followed by the addition of 0.2 mM IPTG after approximately 15 minutes of constant shaking.

Protein Purification Coupled to the FPLC system

Nickel Sepharose Fast Flow Affinity Chromatography for His_6-Tagged Proteins

The cells were harvested at 6000 rpm for 10 minutes. For a 1 liter culture, the cell pellet was resuspended in 25 mL buffer A containing 25 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole. Once resuspended, the cells were lysed by sonication in a 50 mL centrifuge tube for 3 x 1 minute (duty cycle 50%, output 7) followed by centrifugation at 11,000 rpm for 30 minutes. Once pelleted, the cell lysate was filtered through a 1.5 micron filter (Whatman) prior to loading onto a pre-equilibrated 7 mL Nickel Sepharose 6 Fast Flow (GE Healthcare) column at a flow rate of 3 mL/min. The column was washed with 7 column volumes (CV) of Ni buffer A to remove unbound sample. The protein of interest was eluted in 3 mL fractions using Ni buffer B containing 300 mM imidazole during a 4 CV gradient length followed by 60 mL of elution with 100% buffer B. The protein typically eluted in 35-40 mL total volume.

Amylose Affinity Chromatography for MBP Fusion Proteins

For a 1 liter culture, the cell pellet was resuspended in 25 mL buffer A containing 25 mM Tris pH 7.5, 500 mM NaCl following centrifugation for 10 min at 6000 rpm. Once resuspended, the cells were lysed by sonication in a 50 mL centrifuge tube for 3 x 1 minute (duty cycle 50%, output 7) followed by centrifugation at 11,000 rpm for 30 minutes. Once pelleted, the cell lysate was filtered through a 1.5 micron filter (Whatman) prior to loading onto a pre-equilibrated 10 mL Amylose (New England BioLabs) column at a flow rate of 2 mL/min. The column was washed with 10 CV of amylose buffer A to remove unbound sample. The protein of interest was eluted in 3 mL fractions using amylose buffer B

containing 30 mM maltose during a 3 CV gradient length and 40 mL gradient delay. The protein typically eluted within 20-25 mL total volume.

Glutathione Sepharose 4 Fast Flow Affinity Chromatography for GST fusion proteins

Glutathione Sepharose 4 fast flow affinity chromatography is used to purify glutathione S-transferase (GST) fusion proteins. This resin contains glutathione-agarose beads that bind GST-tagged proteins which can be eluted from the resin using free glutathione. The cell pellet harvested from a 1L culture was resuspended in 25 mL GST buffer A containing 20 mM NaH_2PO_4 pH 7.2, 500 mM NaCl. The cells were then disrupted by sonication on ice for 2 x 30 seconds (duty cycle 50, output control 7) followed by centrifugation at 11,000 rpm for 30 min in a 50 mL centrifuge tube. The supernatant was filtered through a 1.5 micron filter and then loaded onto a pre-equilibrated 10 mL Glutathione Sepharose 4 Fast Flow (GE Healthcare) column at a flow rate of 2 mL/min. The column was washed with 10 CV buffer A to remove unbound sample. The GST-tagged protein was eluted in 3 mL fractions using a gradient length of 3 CV and delay of 90 mL. The protein typically eluted in 40-45 mL total volume.

Hi-Load Superdex 75/200 Gel Filtration Chromatography

Gel filtration chromatography can be used as a final purification for a protein of interest in order to remove impurities such as aggregated protein or free tag like thioredoxin or maltose binding protein which co-elute during affinity chromatography with the desired protein. The fractions from the affinity column were combined and the protein was concentrated to approximately 5 mL (~ 10 mg/mL final concentration) using an Amicon centrifugal concentrator. The concentrated protein was then injected into a 10 mL sample loop and loaded onto the 100 mL Superdex 75 (200 for larger proteins) filtration column pre-

equilibrated overnight with 20 mM MOPS pH 7.0, 400 mM NaCl. The column was washed with the MOPS buffer at 1 mL/min collecting 2 mL fractions. To determine which peak is the protein of interest, a set of molecular mass standards including cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), BSA (66 kDa), and β-amylase (200 kDa) (Sigma Aldrich) were eluted using the same buffer conditions as the protein of interest. Approximately 100 µg of each standard was present in the standard mix. The void volume (V_0) and total column volume (V_t) were determined using 2 mg blue dextran (2000 kDa) and 0.025% phenol red (0.4 kDa) respectively. The log of the molecular mass of each protein standard was then plotted against the retention time, $[(V_e-V_0)/(V_t-V_0)]$, in the column, generating a standard curve that was used to determine the apparent molecular mass of the protein of interest. Based on the theoretical and apparent molecular masses, the fractions pertaining to the protein of interest can then be combined. In most cases, the first peak eluting within the void volume is consistent with the GroEL/GroES chaperone complex. If monomeric, the target protein will be well separated from the chaperone complex peak. However, if present in a higher oligomeric state (i.e. trimer), the protein will be eluted rather close to the chaperone peak. To ensure complete elimination of potential impurities, the fractions pertaining to the chaperone complex and target protein must be analyzed for enzymatic activity.

Protein Property Analysis using Gel Filtration Chromatography Using HPLC

The solubility and oligomeric state of a protein of interest was monitored by gel filtration coupled to the HPLC. The protein of interest was concentrated sequentially to 2x, 4x, 8x the initial concentration and 100 uL of the protein at each concentration was injected onto a G3000SW_{XL} (Tosoh Bioscience) gel filtration column pre-equilibrated with running buffer containing 20 mM MOPS, 400 mM NaCl, pH 7.0. The column was washed with

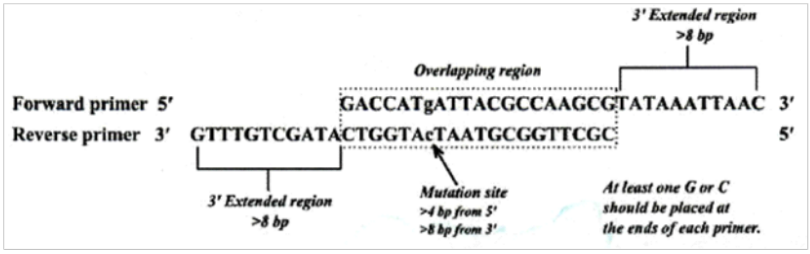
running buffer with a flow rate of 1 mL/min and the chromatogram was recorded by UV absorbance at 280 nm. The molecular weight of the target protein can be determined through comparison to known protein standards. If the threshold of protein solubility was not reached, then there would be one peak area that increased in proportion to the protein concentration. However, if the threshold was reached, then there would be a point at which the peak area would not be proportionally increased as the protein concentration increased.

Site-Directed Mutational Analysis

Design of Primers and Polymerase Chain Reaction

The forward and reverse primers containing the overlapping mutation site were designed as described previously (104,230) (Fig 23A). The PCR reaction was conducted by mixing the following components: 10X Pfu buffer, 5µL; 10 mM dNTP mix (New England BioLabs), 1 µL; 50-100 ng plasmid DNA; forward and reverse primers (25 µM), 1.5 µL each; Pfu Turbo DNA polymerase (Stratagene), 1 µL; water to a final volume of 50µL. The PCR cycle utilized for these reactions is described in Figure 23B. Once the PCR reaction was completed, a 10 µL aliquot was examined by agarose gel electrophoresis using a 1% agarose gel. Once the PCR product was identified on the gel, 10 µL of the PCR product was digested with restriction enzyme Dpn I (New England BioLabs) at 37ºC for 1 hour. A 5 µL aliquot of the Dpn I digested product was used to transform 50 µL XL10 gold competent cells.

A.



B.

Step	Number of Cycles	Temperature (°C)	Time
Initialization	1	94	3 min
Denaturing	16	94	1 min
Annealing		52	1 min
Elongation		68	1 min/kilobase
Final Elongation	1	68	1 hour

Figure 23. Guidelines for site-directed mutagenesis polymerase chain reaction. **(A)** Strategy for design of site directed mutagenesis primers (230). **(B)** Thermocycler program for site-directed mutagenesis PCR.

Small Scale Expression and Purification of Mutant Enzymes

The mutant plasmid was chemically transformed into the *E. coli* protein expression cell line (BL21 (DE3), Origami B (DE3), Origami B^{chaperone} (DE3)) followed by plating on LB-agar plates with the appropriate antibiotics. It is important to note that each of these cell lines has varying antibiotic resistance as follow: BL21 (DE3), none; Origami B (DE3), tetracycline and kanamycin; Origami B^{chaperone} (DE3), tetracycline, kanamycin, and chloramphenicol. Once colonies were formed following overnight incubation at 37°C, a single colony was used to inoculate 3 mL LB media with the appropriate antibiotics. The culture was grown overnight at 37°C with constant shaking at 250 rpm. A 1 mL aliquot of the saturated overnight culture was then used to inoculate 100 mL sterile LB media with the appropriate antibiotics. The culture was grown with constant shaking at 250 rpm at 37°C

until the $O.D._{600}$ reaches 0.6-0.8. The temperature was then reduced to 22°C followed by the addition of 0.2 mM IPTG to induce protein expression. Take note, 1 mg/mL L-arabinose must be added prior to IPTG to induce GroEL/GroES chaperone expression for the Origami $B^{chaperone}$ cells. The cells were then allowed to grow overnight at 22°C with constant shaking at 250 rpm followed by harvesting with centrifugation at 6000 rpm for 30 minutes. The cell pellet was resuspended in buffer A corresponding to the column needed for purification (i.e. nickel, amylose, heparin, GST, or DEAE). The resuspended cells were then sonicated 2 x 30 seconds (output control 4) and the cell lysate was spun at 11,000 rpm for 20 min. The cell lysate was filtered using a 0.45 µm syringe filter and the supernatant was loaded onto a pre-equilibrated 500 µL column containing the necessary resin. The column was washed two times with 3-4 mL buffer A followed by elution with 1 mL buffer B corresponding to the resin used.

Sulfotransferase Activity Analysis of Mutant Enzymes

The sulfotransferase activity was determined by incubating approximately 0.3-0.8 µg of purified wild type or mutant protein with either 5 µg of completely de-*O*-sulfated *N*-sulfated heparin (from Neoparin) or 1 µg *N*-sulfo heparosan and 1-5 × 10^5 cpm of [^{35}S]PAPS in 100 µL of a buffer containing 50 mM MES, pH 7.0, 10 mM $MnCl_2$, 5 mM $MgCl_2$, and 1% Triton X-100. The reaction was incubated at 37°C for 30 min and quenched by the addition of UPAS buffer containing 50 mM NaOAc pH 5.0, 250 mM NaCl, 3 M urea, 1 mM EDTA and 0.01% Triton X-100. The sample was then loaded onto a 200 µL DEAE-Sepharose (Sigma-Aldrich) column to purify the [^{35}S]heparan sulfate. The column was washed 4 times with UPAS buffer followed by 3 times with a buffer containing 250 mM NaCl and 0.001%

Triton X-100. The purified [^{35}S]HS was eluted with 1 M NaCl and 0.001% Triton X-100. The quantity of [^{35}S]HS was determined by liquid scintillation counting.

Large Scale Preparation and Purification of K5 Polysaccharide

The *E.coli* strain NCDC Bi 8337-41 (ATCC# 23506) was utilized for production of the K5 bacterial polysaccharide. A 100 mL culture containing sterile LB medium was inoculated with a glycerol cell stock of K5P bacterial polysaccharide and grown at 37°C for approximately 8 hours. A 13 mL aliquot from the 100 mL culture was then used to inoculate one L cultures for 6 L and grown for 16-18 hours at 37°C. The cells were then harvested by centrifugation at 7000 rpm for 30 min. The supernatant was filtered through a 1.5 µM filter (Whatman) and the pH of the solution was adjusted 5. Once the pH was adjusted, the supernatant was mixed at 1:1 ratio with DEAE column buffer A containing 20 mM sodium acetate pH 5.0 and 50 mM NaCl. The solution was loaded onto a 100 mL DEAE Sepharose fast flow (GE Healthcare) column pre-equilibrated with water followed by 1 L filtered buffer A at a flow rate of 10 mL/min. After the material was completely loaded, the column was washed with 1L filtered A before switching to elution buffer B containing 1M NaCl. Buffer B was loaded onto the column up to a volume of approximately 200 mL and then the flow rate was stopped for 1 hour before eluting the K5P. The K5P was eluted with approximately 400 mL buffer B or until eluent was clear. The collected K5P eluent was ethanol precipitated overnight at -20°C by mixing the eluent with an equal volume of 100% reagent ethanol. The precipitate was centrifuged at 7000 rpm for 30 minutes and resuspended in water before splitting equally into two 50 mL centrifuge tubes. A final volume of 50% saturated ammonium sulfate was added to the precipitated K5P and incubated on ice for 15 minutes. The gel like precipitate was collected by spinning at 6000 rpm for 30 minutes followed by

removing the supernatant from the gel like precipitate. This process was repeated using a final volume of 60% saturated ammonium sulfate. The gel precipitate was resuspended in 10 mL water and dialyzed overnight in water with MWCO 12-14,000 membrane (Spectrapor), leaving 2 times the amount of headspace for expansion. The dialyzed sample was recovered and loaded onto a 20 mL DEAE Sepharose fast flow (GE Healthcare) column connected to the FPLC pre-equilibrated with buffer containing 20 mM sodium acetate, 50 mM NaCl pH 5.0 at a flow rate of 2 mL/min. The method was first calibrated by mixing a small volume of K5P (~5-10 mL) with a trace of 150 µL [^{3}H]K5P to monitor elution of the K5. Every other fraction eluted from the DEAE column was counted for radioactivity to check for [^{3}H]K5P. The radioactive K5 eluted just before the protein peak appearing on the FPLC chromatogram. The polysaccharide was eluted in 35 mL fractions using buffer B containing 1M NaCl with a gradient length of 10 CV. The eluted K5P was (1:1) phenol/chloroform extracted by adding 1-2 mL phenol/chloroform and rotating overnight at 4°C. The sample was centrifuged at 6000 rpm for 30 minutes and the supernatant collected followed by ethanol precipitation overnight at -20°C. The sample was pelleted at 6000 rpm for 30 minutes and resuspended in DEAE column buffer A before loading onto the pre-equilibrated DEAE column. The same method was repeated eluting with buffer B containing 20 mM sodium acetate pH 5 and 1 M NaCl and collecting 4 mL fractions over a gradient length of 10 CV. The K5P again eluted in a total volume of 35 mL. The pooled K5P was dialyzed overnight against water in MWCO 12-14,000 membrane. The sample was dried down and the tubes were weighed to determine the approximate amount of K5P.

Modification of Heparan Sulfate Polysaccharides Using PAPS Regeneration System

The PAPS regeneration system can be utilized to overcome product inhibition by PAP when attempting to sulfate polysaccharides. This system relies on aryl sulfotransferase IV (Ast-IV) and a high concentration of *p*-nitrophenyl sulfate (PNPS) to maintain a continuous supply of PAPS. To enzymatically sulfate heparan sulfate polysaccharides, the K5 (~ 10 mg) was mixed with purified 5 mL Ast-IV (1 mg/mL), 40 µM PAP, 1 mL of 100 mM PNPS, and 50 mM MES pH 7.0, 0.5% Triton X-100 to a final volume of 40 mL. The yellow-colored reaction mixture was incubated for 15 min at RT with constant nutation to initiate PAPS production followed by the addition of 5 mL (~5 mg) heparan sulfate sulfotransferase. The reaction was incubated overnight at RT with constant nutation. When working with multimilligram quantities of K5P, multiple reactions must be prepared assuming 10 mg K5P per reaction mixture. Following incubation, each reaction mixture was centrifuged at 6000 rpm for 20 minutes to pellet any precipitated enzyme. The material was loaded onto a 3 mL DEAE Sepharose fast flow column pre-equilibrated with 5 CV water followed by 10 CV buffer A containing 20 mM Tris pH 7.4, 250 mM NaCl. The DEAE column was washed with 5 CV buffer A followed by elution with 1 CV buffer B containing 1 M NaCl. The eluted material was dialyzed against 5 mM ammonium bicarbonate using MWCO 3500 membrane (Spectrapor) overnight at 4°C, dried down to completion and resuspended in distilled water.

Disaccharide Analysis of Heparan Sulfate Polysaccharides

Enzymatic Degradation Using Heparin Lyases

Cleavage with heparin lyases results in the generation of a $\Delta^{4,5}$-unsaturated uronic acid at the non-reducing end, which is easily detectable at a UV absorbance of 232 nm. With

that said, the polysaccharides were digested to disaccharides using a mixture of the heparin lyases. Approximately 10-20 µg modified (or unmodified) K5 polysaccharide was incubated with 200 µL 50 mM Na_2HPO4 pH 7.0 and 10 µL of each heparin lyase (I, II, and III) at 37°C overnight. The reaction was boiled for 2 min at 100°C to terminate the reaction followed by centrifugation at 13,000 rpm for 2 minutes. A 100 µL aliquot of the sample was then mixed with 130 mM tetrabutylammonium dihydrogenphosphate (TBA) and 100 µL reverse phase ion pairing (RPIP) buffer A containing 38 mM $NH_4H_2PO_4$, 2 mM H_3PO_4, 1 mM TBA. The mixture was then loaded onto a RPIP C_{18} column (Vydac 218TP, 5 µm, Cat#218TP 54) connected to the HPLC at a flow rate of 0.5 mL/min with UV detection at 232 nm. This column was pre-equilibrated with 14% buffer B containing 38 mM $NH_4H_2PO_4$, 2 mM H_3PO_4, 1 mM TBA, and 50% CH_3CN. The disaccharides were eluted from the column by increasing the percentage of buffer B from 14% to 30% at 45 min, 39% at 60 min, and then 100% at 85 min. The composition of the disaccharides from the injected sample was then determined by comparison to the elution times of a set of disaccharide standards (Seikagaku Corporation 400576 Unsaturated Heparan/Heparin disaccharide mixture H mix). For the standard mix, 20 µL of 0.1 mM standard mix solution was mixed with 130 mM TBA and 180 µL RPIP buffer A and analyzed using RPIP-HPLC analysis using the same buffer conditions. Please note, if analyzing radiolabeled material, at least 10,000 cpm of [^{35}S] or [^{3}H] HS must be loaded onto the RPIP column to generate sufficient results.

Chemical Degradation Using Low pH 1.5 Nitrous Acid

Low pH nitrous degradation at pH 1.5 is utilized to cleave heparan sulfate polysaccharides after the *N*-sulfo glucosamine residue. This cleavage will maintain the configuration of the uronic acid. The dried [^{35}S]-radiolabeled HS sample was resuspended in

60 µL water, using 1/3 of the material for disaccharide analysis. Approximately 10 µg of cold heparan sulfate must be added to the radiolabeled sample as a cold carrier. A 40 µL aliquot of 1:1 ratio of 0.5 M H_2SO_4:0.5 M $Ba(NO_2)_2$ was added to the HS material. The H_2SO_4:$Ba(NO_2)_2$ mixture must first be centrifuged at 13,000 rpm for 1 min and the supernatant added to the HS. The reaction is incubated on ice for 30 min followed by the addition of 20 µL 7:3 (1M Na_2CO_3:1M $NaHCO_3$) pH 9.5 to terminate the reaction. A 20 µL aliquot of 0.5 M sodium borohydride in 0.1 M NaOH was added to the sample and incubated for 30 min at 50°C. The reaction is then cooled to room temp before the addition of 20 µL of 10 M acetic acid to stop the reduction and release the hydrogen gas. The sample must be vortexed once the acetic acid is added. Phenol red (1 µL of 5% phenol red) was then added as an indicator. If the color of the solution was yellow, then stepwise addition of 10 M NaOH was completed until the color turns light red. The sample was then loaded onto a pre-equilibrated P-2 (BioRad) column using 0.1 M ammonium bicarbonate for elution. The material was collected in 1 mL fractions using a fraction collector. The [^{35}S]-labeled disaccharide products were monitored by radioactive counting of each P-2 fraction, collecting and pooling the eluted material. The disaccharide product was dried down using the Speedvac followed by resuspension in ddH_2O. A 50 µL aliquot (~10-20,000 cpm) of the sample was then mixed with 120 mM tetrabutylammonium dihydrogenphosphate (TBA) and 100 µL ddH_2O to a final volume of 250 µL. The mixture was then loaded onto a RPIP C_{18} column (Vydac 218TP, 5 µm, Cat#218TP 54) connected to the HPLC at a flow rate of 0.5 mL/min with UV lamp setting at 232 nm. Buffer A consisted of 9.5 mM $NH_4H_2PO_4$, 0.5 mM H_3PO_4, 1 mM TBA. The column was pre-equilibrated with 14% buffer B containing 9.5 mM $NH_4H_2PO_4$, 0.5 mM H_3PO_4, 1 mM TBA, and 30% CH_3CN. The disaccharide products

were eluted from the column by increasing the percentage of buffer B from 14% to 30% at 45 min, 39% at 60 min, and then 100% at 85 min. The composition of the disaccharides from the injected sample was then determined by coelution with appropriate disaccharide standards (145).

Isolation of Chicken 2OST DNA

The truncated chicken 2OST cDNA consisting of E52-N356 was cloned by fusing two expressed sequence tags cDNA clones ChEST584h10 and ChEST850d20, which were purchased from ARK-Genomics (Roslin Institute, Midlothian, UK). The *N*-terminal fragment of 2OST (E52-F198) was amplified from the EST clone ChEST584h10 by PCR using two primers: ATAGAACAACGTCATACAGCAGATGGCC (O785) and CAAAGGTCTTCTTATCCCCCTG (O791). An *Nco*I site was then introduced to the *N*-terminal fragment by PCR using primer ATTAATTACCATGGAGAGATAGAACAACGTCATA (O781, where the *Nco*I site is underlined) and primer O791. The *N*-terminal fragment was digested with *Nco*I prior to the ligation with the *C*-terminus as described below. The *C*-terminal fragment of 2OST (D199-N356) was amplified from the EST clone ChEST850d20 by PCR using two primers: pATGAATGTGTGGCAGCTGGA (O790, containing 5'-phosphate) and ATTAAATAAAGCTTTTCTTAGTTTGATTTGGGGT (O781, where the *Hind* III site is underlined). The *C*-terminal fragment was then digested with *Hind* III. Both *N*-terminal and *C*-terminal fragments were ligated by a rapid ligation kit (Roche). The truncated chicken 2OST cDNA (E52-N356) was then amplified from ligated *N*- and *C*-terminal fragments by PCR using primers O780 and O781 and cloned into the PET-32b vector (Novagen) using

*Nco*I and *Hind* III sites. The truncated chicken 2OST cDNA was confirmed by sequencing analysis (University of North Carolina at Chapel Hill Genome Analysis Facility).

Cloning, Expression and Purification of Chicken 2OST

The maltose binding protein (MBP)-2OST fusion protein was prepared using a modified pMAL-c2x vector (New England Biolabs). The amino acid sequence of MBP was truncated at Asn-367 and contained the mutation E359A (59). The linker region encodes three alanine residues (A368-A370) and contained a NotI site for cloning. The catalytic domain of chicken 2OST (D69-N356) was cloned into the vector by using the NotI and BamHI sites. The MBP-2OST protein was expressed in Origami B (DE3) cells (Novagen). Cells were grown in a shaker at 37°C in sterile LB medium and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside once the temperature of the cell culture was drop to 18°C. Cells were pelleted at 6000 rpm for 10 min, resuspended in 25 mM Tris pH 7.5, 500 mM NaCl, and 1 mM DTT, and then lysed by sonication 3 x 1 min (output 7, duty cycle 50). MBP-2OST was bound to amylose resin (New England Biolabs), eluted with 40 mM maltose, and then loaded onto a HR16/60 Superdex 200 (Amersham) column pre-equilibrated in buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT, 40 mM maltose. The purified protein was then dialyzed overnight against 25 mM Tris pH 7.5, 75 mM NaCl, 5 mM maltose, and 1 mM DTT. PAP was added to 1 mM concentration. The sample was concentrated to 19 mg/mL, followed by the addition of more PAP for a final concentration of 4 mM.

Crystallization of Chicken 2OST

Crystals of MBP-2OST were formed by using the sitting drop vapor diffusion technique at 4°C. The sitting drop contained 1 µL of protein solution, mixed with 1 µL of

reservoir solution consisting of 100 mM BisTris-propane pH 6.0, 1.7 M ammonium citrate, and 10 mM phenol. For data collection, crystals were transferred to 100 mM BisTris pH 6.0, 1.8 M ammonium citrate, 1 mM phenol, 25 mM Tris pH 7.5, 75 mM NaCl, 5 mM maltose, 1 mM PAP and 12% ethylene glycol for 45 seconds and then were flash frozen in liquid nitrogen. Diffraction data were collected on a Rigaku 007HF generated equipped with VariMaxHF mirrors and a Saturn92 CCD detector. For the first crystal, data were collected and processed to 2.85Å resolution by using HKL2000. The program MOLREP was used to solve the phase problem using molecular replacement with residues 1-370 of MBP from the MBP-RACK1 crystal structure as the starting model. The search found 1 MBP molecule in the asymmetric unit. To determine the position of 2OST, a search model was created from the crystal structure of mouse EST (Protein Data Bank ID code 1AQU) consisting of residues from conserved features found in both EST and NST or 3OSTs (D39-Y62, R78-E83, I104-I146, L189-F210, and P269-K285). The rest of the 2OST model was built through iterative cycles of model building in O, refinement in CNS, and density modification by using solvent flipping that greatly improved the quality of the electron density maps. A second crystal structure was used to collect a dataset at 2.65 Å. This dataset was used for the final rounds of refinement and model building, maintaining the same set of reflections for the free R calculations as the first out to 2.85 Å. The final model consists of all residues in MBP 1-370 and residues D69-K354 in 2OST with 95.6% of residues in the 98% favored region and 99.7% in the 99.8% allowed region, as determined by MOLPROBITY.

Preparation of Chicken 2OST Mutant Plasmids

The 2OST mutants were prepared using 2OST-pMAL-c2x plasmid as the template and a modified method from the Stratagene QuickChange mutagenesis protocol (230). The

mutagenesis primers were synthesized by Invitrogen. The resultant constructs were sequenced to confirm the expected mutation (University of North Carolina at Chapel Hill Genome Sequencing Facility).

Expression and Purification of 2OST Mutant Proteins

Origami B (DE3) cells were transformed with the 2OST mutant protein expression vector using the method listed above. A single colony was inoculated into 3 mL LB medium supplemented with 50 µg/mL carbenicillin, 12.5 µg/mL tetracycline, and 15 µg/mL kanamycin and grown overnight at 37°C with constant shaking. A 1 mL aliquot of the saturated overnight culture was added to 100 mL LB supplemented with the same antibiotics. The culture was grown at 37°C for 3-4 hours until the $O.D._{600}$ reached 0.6-0.8. The temperature of the shaker was then dropped to 18°C to cool down the culture. The protein expression was induced with 0.5 mM isopropyl-β-$_D$-galactopyranoside with constant shaking for 18-20 hours at 18°C. The cells were harvested by centrifugation at 6000 rpm for 10 min in a 50 mL centrifuge tube. The cell pellet was resuspended in 8 mL buffer A containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT. The cells were lysed by sonication in a small sonication bottle on ice 2 x 45 seconds (output 4, duty cycle 50) followed by centrifugation at 11,000 rpm for 20 min. The cell lysate was filtered through a 0.45 µm syringe filter (Millipore). The filtered lysate was loaded onto a disposable poly-prep column (Bio-Rad) containing 500 µL amylose-agarose (New England Biolabs) resin which has been pre-equilibrated with buffer A. The protein was loaded onto the amylose resin, washed with 2 CV buffer A, and then eluted with 1 M buffer B containing 40 mM maltose. A small aliquot of the 2OST mutant proteins were utilized to analyze for sulfotransferase activity as described under the section entitled *Sulfotransferase Activity Analysis of Mutant Enzymes.*

Gel Filtration Chromatography of 2OST Wild Type and 2OST V332STP

To determine the oligomeric state of 2OST WT and 2OST V332STP, approximately 7 mg of amylose-agarose purified 2OST WT or 2OST V332STP was fractionated by gel filtration chromatography on a Superdex 200 HiLoad 16/60 column that was equilibrated with a buffer containing 20 mM MOPS pH 7.0, 400 mM NaCl, and 40 mM maltose at a flow rate of 1 mL/min. Fractions of 2 mL each were collected. The eluate was monitored by UV absorbance at 280 nm. The apparent molecular mass of 2OST WT or V332STP was determined by comparison with the elution positions of molecular mass standards, including cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), BSA (66 kDa), and β-amylase (200 kDa) (Sigma Aldrich). The standard mix contained approximately 100 µg of each standard listed. The void volume (V_0) and total column volume (V_t) were determined by using 2 mg blue dextran (2,000 kDa) and 0.025% phenol red (0.4 kDa) respectively. The protein standards were eluted using the same buffer conditions as for the 2OST protein.

Determination of the Substrate Specificity of 2OST R189A

Since *N*-sulfo heparosan contains no IdoUA unit, it was incubated with recombinant C_5-epi to synthesize the IdoUA unit followed by 2OST modification. For a 100 µL reaction, *N*-sulfo heparosan (9 µg) was incubated with C_5-epi (7 µg) in a buffer containing 50 mM MES (pH 7.0), 1% Triton X-100, and 2 mM $CaCl_2$ for 1 hour at 37°C. To this reaction, 10 µg of 2OST R189A or 2OST WT and $1.5x10^6$ cpm of [^{35}S]PAPS were added and incubated at 37° for 1 hour. The 2-*O*-[^{35}S]sulfo group in the polysaccharide was purified by a DEAE column. The position of the 2-*O*-[^{35}S]sulfo group in the polysaccharide was determined by disaccharide analysis. Briefly, the resultant ^{35}S-labeled polysaccharide was degraded to disaccharides with nitrous acid at pH 1.5 followed by reduction with sodium borohydride.

The disaccharide products were then analyzed by reverse phase ion-pairing HPLC. The identities of the resultant ^{35}S-labeled disaccharides were determined by coelution with appropriate disaccharide standards (145).

Preparation of Recombinant C_5-Epimerase

Human C_5-epi was expressed in Origami-B DE3 cells (Novagen), which contain pGro7 (Takara, Japan) plasmid expressing chaperonin proteins GroEL and GroES of *E. coli*. Different *N*-terminal truncated forms of C_5-epi were cloned into pMalc2x vector (New England BioLab) to form maltose-binding protein (MBP)/Epi fusion proteins. Several *N*-truncation points were prepared at Asp-34, Glu-53, Gln-72, Asn-92, Trp-203 and Phe-257. The transformed cells were grown in LB medium supplemented with 12.5 μg/mL tetracycline, 15 μg/mL kanamycin, 40 μg/mL chloramphenicol and 50 μg/mL carbenicillin at 37 ^{o}C. The temperature was decreased to 22 ^{o}C when the A_{600} reached 0.4-0.8 followed by addition of L-arabinose (1 mg/mL) and isopropylthiogalactopyranoside (0.5 mM) to induce the expression of chaperone and C_5-epi proteins, respectively. After shaking overnight, the cells were pelleted and lysed by sonication. The expressed proteins were then purified using an amylose-agarose (New England BioLabs) column following the protocol provided by the manufacturer. The protein product was analyzed by 10% SDS-PAGE stained with coomassie blue.

Coupling 2OST and C_5-epi to Determine the Activity of C_5-epi

For a 100 μL total reaction, wild type C_5-epi (1.5 μg) was incubated with 2 μg *N*-sulfo heparosan in the buffer containing 50 mM MES (pH 7.0), Triton X-100 and 1 mM $CaCl_2$ at 37 ^{o}C for 30 minutes. The mixture was heated at 100^{o}C for 5 min and then cooled down to room temperature. After centrifugation to discard the insoluble matters, 0.6 μg

2OST Y94I, 0.01 mg bovine serum albumin and [^{35}S]PAPS (5-7 × 10^5 cpm) were added to the reaction followed by incubation at 37^{o}C for another 5 min. The resultant ^{35}S-labeled polysaccharide was purified by DEAE chromatography. To determine the linear range of this assay, a dose response curve was also generated using wild type C_5-epi (0-3.3 μg) and 2OST Y94I mutant (0.6 μg). The activities of various C_5-epi mutants (1.5 μg) were also tested using this protocol.

Disaccharide Analysis of the Polysaccharide Modified by C_5-epi and 2OST WT

The ^{35}S-labeled polysaccharides were prepared as described above except that 2OST WT was used. Here we scaled up the reaction volume to 3 ml. Briefly, After a 30-min incubation of ***N***-sulfoheparosan with C_5-epi or its mutants, the reaction was terminated by heating at 100 °C for 5 min. Then 2OST WT, bovine serum albumin, and [^{35}S]PAPS were added to the supernatant after the reaction mixture was spun down and incubated at 37 °C for an additional 5 min. After DEAE column purification, dialysis against deionized water, and drying by speed vacuum, the polysaccharide products were subjected to nitrous acid degradation at pH 1.5 (29). The disaccharides were separated by a BioGel P-2 column, which was equilibrated with 0.1 M ammonium bicarbonate (27, 30). The resultant disaccharides were resolved on a C_{18} column (0.46 × 24 cm; Vydac) under the reverse-phase ion-pairing (RPIP)-HPLC conditions. Here, the column was eluted with a stepwise gradient of acetonitrile containing 0.5 mM H_3PO_4, 9.5 mM $NH_4H_2PO_4$, and 1 mM tetrabutylammonium phosphate monobasic (Sigma). The column was washed with 5.2% acetonitrile for 30 min first, followed by 15 min of 9% acetonitrile elution and then 60 min of 11.7% acetonitrile elution. The identities of the disaccharides were determined by coeluting with authentic disaccharide standards (176).

Preparation of 6OST-3 Mutant Plasmids

The 6OST-3 mutants were prepared using 6OST-3-pET21b plasmid as the template and a modified method from the Stratagene QuickChange mutagenesis protocol (230). The mutagenesis primers were synthesized by Invitrogen. The resultant constructs were sequenced to confirm the expected mutation (University of North Carolina at Chapel Hill Genome Sequencing Facility).

Small Scale Expression and Purification of 6OST-3 Mutant Proteins

Origami-B^{chaperone} cells were transformed with the mutant 6OST-3-pET21b plasmid. A single colony was inoculated into 3 mL Luria-Bertani (LB) media supplemented with 12.5 µg/mL tetracycline, 20 µg/mL chloramphenicol, 15 µg/mL kanamycin, and 50 µg/mL carbenicillin and allowed to grow overnight at 37°C. One mL of the overnight culture was used to inoculate 100 mL LB supplemented with the aforementioned antibiotics at their corresponding concentrations. The culture was shaken at 250 rpm at 37°C until the OD_{600} reached between 0.4 and 0.6. Once the appropriate OD_{600} value was reached, the shaker temperature was reduced to 22° C to allow the culture to cool down. After cooling, 1 mg/mL arabinose was added to induce chaperone expression followed by the addition of 0.2 mM IPTG to induce protein expression. The culture was allowed to shake 18 to 20 hours at 22°C. The cells were harvested by centrifugation for 10 min at 6000 rpm. The cell pellet was resuspended in 8 mL Ni-FF buffer A (30 mM Imidazole, 500 mM NaCl, 25 mM Tris pH 7.5). The resuspended cells were then lysed by sonication for 3 x 45 seconds (output control 7, constant duty cycle) and centrifuged at 14000 rpm for 20 min. Once the spin was completed, the supernatant was filtered through a 0.45 µM syringe filter and loaded onto a 200 µL Ni-FF gravity column pre-equilibrated with 3 mL of Ni-FF buffer A. After the

unbound protein passed through the column, the column was washed with 3 mL buffer A followed by eluting the protein of interest with 1 mL Ni-FF buffer B (300 mM Imidazole, 500 mM NaCl, 25 mM Tris at pH 7.5).

Sulfotransferase Activity Analysis of 6OST-3 Mutants

To test the sulfotransferase activity of the 6OST-3 mutant proteins, a 50 µL reaction was prepared containing 10% Triton X-100, 2.5 µL; 1 M MES pH 7, 2.5 µL; 1 M $MgCl_2$, 0.25 µL; 2 M $MnCl_2$, 0.25 µL, 10 mg/mL BSA, 0.6 µL; 5 µL mutant enzyme, 5 µg *N*-sulfated completely de-O-sulfated heparin, 1 µL; $1x10^7$ cpm powder [^{35}S]PAPS; ddH_2O, 37.9 µL. A negative control without enzyme and a positive control with wild-type 6OST-3 were also included for analysis. The enzymatic reaction was incubated at 37°C for 40 min and loaded onto a DEAE (diethyl aminoethyl) column. The column was washed 2 times with 1 mL UPAS buffer (50 mM NaOAc pH 5.5, 150 mM NaCl, 6 M urea, 1 mM EDTA, 0.1% Triton X-100) and then 3 times with 1 mL 250 mM NaCl solution (50 mM NaOAc, pH 5.5, 250 mM NaCl, 0.01% Triton X-100). The sample was eluted from the column with 1 mL of 1 M NaCl solution (50 mM NaOAc pH 5.5, 1 M NaCl, 0.01% Triton X-100) and 100 µL of eluent was mixed with 2 mL scintillation fluid to determine [^{35}S] radioactivity. The activity level for the mutant protein was then determined using the following calculation: (mutant cpm – negative control cpm)/wild-type cpm. An activity level of less than 10% signified that an amino acid of interest was indeed necessary for activity.

Expression and Purification of Trx-6OST-1

The thioredoxin (Trx)-chicken 6OST-1 fusion protein constructs were created using a pET32(TEV) vector which contains a TEV protease cleavage site for removal of the thioredoxin tag. The 6OST-1-pET32(TEV) plasmid was transformed into Origami B (DE3)

cells. A 100 mL culture of LB supplemented with 12.5 µg/mL tetracycline, 15 µg/mL kanamycin, and 50 µg/mL carbenicillin was inoculated with a glycerol cell stock of 6OST-1-pET32(TEV) and grown overnight at 37ºC with constant shaking at 250 rpm. A 15 mL aliquot of the overnight culture was added to each of the 12L cultures supplemented with the same three antibiotics at the appropriate concentrations. The cultures were grown at 37°C until the $O.D._{600}$ reached 0.6-0.8 following by reduction of the incubator temperature to 22°C. Next, the addition of 0.1 mM IPTG was completed to induce protein expression. The cultures were allowed to shake overnight at 22°C. The cells were pelleted at 6000 rpm for 10 min followed by resuspension in 75 mL Nickel buffer A containing 25 mM Tris pH 7.5, 30 mM imidazole. A 1:100 ratio of phenylmethylsulfonyl fluoride (PMSF) was added to inhibit serine protease digestion. At this point, the purification by nickel column was completed as described in *Nickel Sepharose Fast Flow Affinity Chromatography for His_6-Tagged Proteins*. Once the eluted protein fractions were pooled, PAP was added to a final concentration of 1 mM to keep the protein in solution. The fusion protein was then treated with TEV protease (1:25 digestion) and shaken gently overnight at 4°C. The TEV cleaved 6OST-1 was diluted 3x with heparin buffer A containing 25 mM Tris pH 7.0 and 150 mM NaCl and loaded 5 times onto a 10 mL heparin column coupled to the FPLC at 4 mL/min with a gradient length of 10 CV and delay of 75 mL. The protein was eluted in 4 mL fractions with buffer B containing 25 mM Tris pH 7.0 and 1 M NaCl. The eluted protein from each round of heparin purification was pooled and concentrated to 10 mL final volume. The heparin purified 6OST-1 was loaded twice (5 mL per injection) onto a pre-equilibrated Superdex 75 column at a flow rate of 1 mL/min and eluted in 1 mL fractions with a buffer containing 20 mM

MOPS pH 7.0 and 400 mM NaCl. The eluted 6OST-1 protein fractions from each GPC run were pooled and stored at -80°C.

Chapter III

Structurally Guided Mutational Analysis of Heparan Sulfate 2-*O*-Sulfotransferase

Heparan sulfate 2-*O*-sulfation is a critical modification within the HS biosynthetic pathway. This critical modification is carried out by heparan sulfate 2-*O*-sulfotransferase (2OST). The product is involved in mediating signal transduction pathways for fine structure formation in different species. It has been found to be important for kidney development in mice as well as limb bud formation in chicken (100,102,103). Understanding the substrate recognition mechanism of this enzyme is essential for determining the role of 2OST in embryonic development as well as for the enzymatic approach to obtaining biologically relevant HS structures. Preliminary substrate specificity and mutagenesis studies have gained some insight into the molecular mechanism of 2OST (65,99,104).

In this chapter, we present the homotrimeric crystal structure for 2OST in complex with 3'-phosphoadenosine 5'-phosphate (PAP). Structural based mutational analysis along the 2OST active site identified several amino acid residues that are responsible for the substrate binding, catalysis, and trimer formation. Interestingly, we identified three amino acid residues that were responsible for redirecting the substrate specificity of 2OST. The mutant 2OST R189A only transferred sulfates to GlcUA moieties within the HS

polysaccharide whereas the mutants Y94A and H106A preferentially transferred sulfates to the IdoUA units (59).

Crystallization of 2OST

Over the last few years, several attempts to crystallize 2OST were made but proved unsuccessful. Initially, a construct of Chinese hamster 2OST with an *N*-terminal histidine$_6$ tag ($^{N'\text{-His}}$2OST) was prepared for purification however this protein had very poor solubility. It was established that the *C*-terminal His$_6$ tag enhanced solubility 10 fold compared to an *N*-terminal His$_6$ tag (126). Although the $^{C'\text{-His}}$2OST protein had increased expression and solubility, the protein was present in more than one oligomeric state making crystallization virtually impossible (126). Ideally, a protein must be present in one single homogeneous population for any hopes of crystallization as it helps to improve crystal formation. Next, a construct expressing 2OST derived from chicken with an *N*-terminal thioredoxin (Trx) tag was prepared, as a Trx tag is known to enhance solubility of a target protein (232). Although the thioredoxin 2OST construct had a much higher level of expression compared to the $^{C'\text{-His}}$2OST, solubility become a major issue especially when attempting to cleave the Trx tag. The final and most promising construct we tested was a chimera of maltose binding protein (MBP) fused to the catalytic domain of 2OST derived from chicken. Chicken was selected because it has a higher body temperature at 42°C, in theory making it more stable and amenable to crystallization (233). This fusion protein was highly expressed, very soluble and existed in a single homogeneous population thus making crystallization a possibility.

For preparing the expression construct, the catalytic domain of chicken 2OST (D69-N356) was cloned into a modified pMAL-c2x vector possessing three alanine residues (A368-A370) in the linker region to reduce flex in the fusion protein to allow for

crystallization. The MBP-2OST fusion protein was expressed in Origami B (DE3) cells and purified by affinity chromatography using an amylose column. From the SDS-PAGE analysis in Figure 24, several bands were present on the gel along with the MBP-2OST protein (70 kDa) including chaperonin proteins (55 kDa) and MBP (42 kDa) that must be eliminated in order to attempt crystallization. The protein utilized for crystallization trials must possess high purity (>90%) therefore any impurities, such as the chaperonin proteins and MBP, that may interfere with crystallization must be removed. A second round of purification was then employed using a Superdex 200 gel filtration column to separate the fusion protein from impurities including aggregated protein and free MBP. Based on the two-step purification, we were able to obtain a highly active MBP-2OST fusion protein with greater than 95% purity.

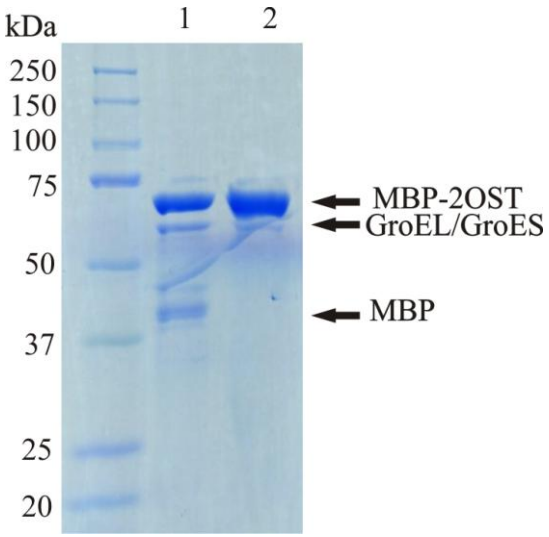


Figure 24. SDS-PAGE analysis of purified wild type MBP-2OST. A small aliquot (5 µg) of MBP-2OST purified by amylose resin (lane 1) and 2OST purified by gel filtration chromatography (lane 2) were analyzed by SDS-PAGE using a 10% Tris-HCl gel stained with Coomassie Blue. The purity of the MBP-2OST is greater than 95% following GPC purification.

Gel filtration chromatography analysis revealed that the fusion protein appeared to behave as a trimer in solution (

Figure **25**). The molecular mass of monomeric MBP-2OST is approximately 70 kDa. A standard curve was generated using the elution volume and molecular mass of known protein standards including β-amylase (200 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome c (12.4 kDa). The void volume (V_o) and total column volume (V_t) were determined by resolving blue dextran (2,000 kDa) and phenol red (0.4 kDa) respectively. A linear plot of log molecular weight against average retention time ($K_{average}$) in the column gave a slope of -3.87±0.25, y-intercept of 2.93±0.09, and r^2 value of 0.99, providing a calculated molecular mass of 210 kDa for the MBP-2OST protein (Fig 23). The calculated molecular mass was consistent with that of a trimer.

A low-resolution crystallographic data set collected from crystals of the catalytic domain of hamster 2OST, lacking a fused MBP molecule, revealed similar trimer formation. The hamster 2OST was generated by cloning the catalytic domain of hamster 2OST (R63-N356) into the pET-21b vector (Novagen) by using NdeI and HindIII sites to form a *C*-terminal $(His)_6$-tagged 2OST expression construct. This dataset provided support that the MBP molecule did not contribute to formation of the trimeric 2OST complex since the hamster 2OST protein did not possess a MBP tag.

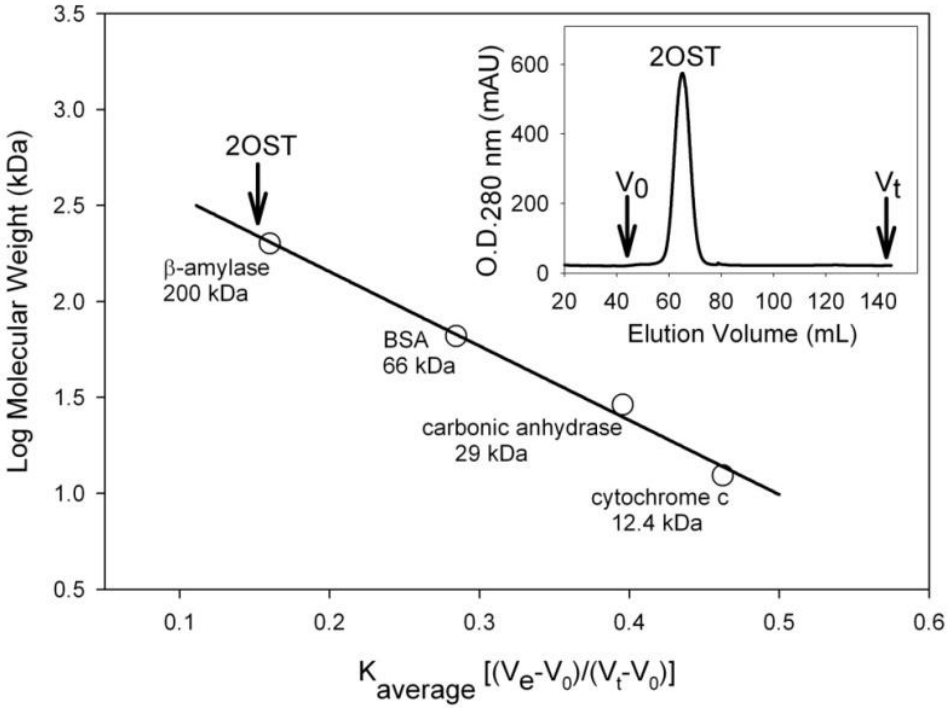


Figure 25. Determination of the apparent molecular weight of 2OST. The molecular weight of the purified 2OST protein was measured by gel filtration chromatography using a Superdex 200 HiLoad 16/60 column. The UV absorbance at 280 nm was plotted against elution volume, revealing a single peak representing the MBP fusion protein (labeled *2OST*), as displayed in the inset. The molecular weight standards used were cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), β-amylase (200 kDa). The void volume (V_o) and total column volume (V_t) were determined by resolving blue dextran (2,000 kDa) and phenol red (0.4 kDa) respectively. The log of the molecular weight for each protein standard was plotted against the average retention time ($K_{average}$) in the column giving a slope of -3.87±0.25, y-intercept of 2.93±0.09, and r^2 value of 0.99. The MBP fusion protein is indicated by an *arrow* labeled *2OST*.

Structural Analysis of 2OST

Comparison to Other Known HS Sulfotransferases

Through a collaborative effort with Dr. Lars Pedersen from NIEHS, the crystal structure of the fusion protein containing MBP and the catalytic domain of chicken 2OST was successfully solved. The crystal structure of MBP-2OST was obtained to 2.65Å in a

binary complex with the inactive sulfo donor, 3'-phosphoadenosine 5'-phosphate (PAP), as seen in Figure 26.

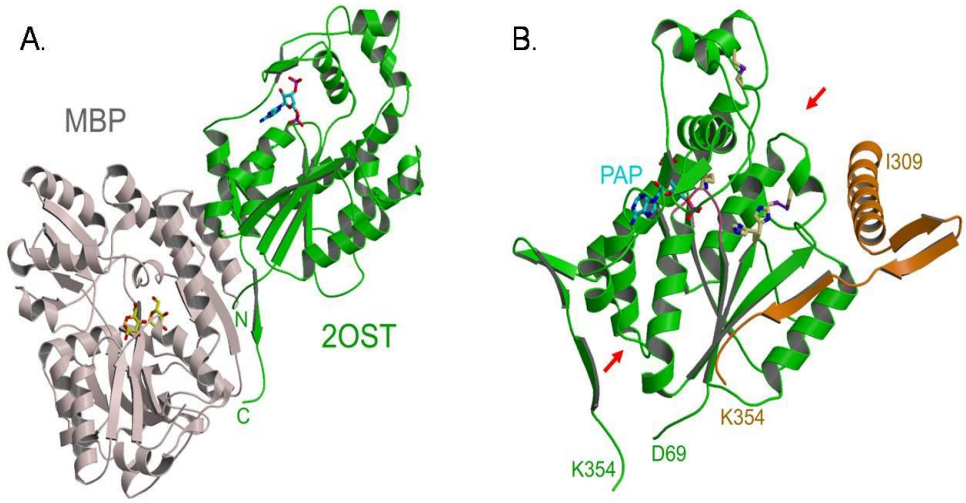


Figure 26. Crystal structure of 2OST. (A) Ribbon diagram of the chimeric construct of MBP (pink) fused to the *N*-terminus of the catalytic domain (Asp69-Lys354) of 2OST (green). A maltose molecule (yellow) is bound to the MBP domain while a PAP molecule (cyan) is bound to the 2OST catalytic domain. *N*- and *C*-terminal ends of the 2OST domain are labeled. **(B)** Ribbon diagram of the 2OST catalytic domain (green) with PAP bound (cyan). The PSB-loop containing Lys83 is colored pink. The two conformations of catalytic base His142 are also shown. Residues Ile309-Lys354 of an adjacent molecule in the trimer are shown in orange and the disulfide bonds are shown in purple. Red arrows indicate the substrate binding cleft. All protein structure figures were created using MOLSCRIPT and Raster 3D (162,234).

The crystal structure revealed that the catalytic domain of 2OST was composed of an α/β motif with a central parallel β-sheet flanked by α-helices on both faces of the sheet (Figure 26A). As seen in Figure 26B, a large open cleft for acceptor carbohydrate substrate binding was found running across one side of the enzyme exposing the active site, similar to the other HS sulfotransferases including the *N*-sulfotransferase domain of *N*-deacetylase/*N*-sulfotransferase (NST), and the 3OST isoforms 1, 3, and 5 for which crystal structures exist (63,79,150,151). Despite their global similarities, the crystal structure for 2OST possessed distinct differences from the other HS sulfotransferases. Sequence alignments of 2OST to NST and 3OST isoforms only demonstrated approximately 12% sequence identity compared

to 28% between NST and the 3OST isoforms. Comparison of 2OST to NST and the 3OSTs revealed the lack of an exterior 3-stranded antiparallel β-sheet containing a disulfide bond and a small loop off the last strand of the central parallel β-sheet where the proposed catalytic base glutamic acid (Glu-184 in 3OST-3) resided. The crystal structure of 2OST exhibited a unique enlongated *C*-terminal helix followed by a 20 amino acid tail extending away from the catalytic domain (Figure 26B and Figure 27).

The crystal structure of 2OST was indeed strikingly different from the other HS sulfotransferases due to the presence of this extended *C*-terminal tail which formed an antiparallel strand arrangement (N345-Y352) with the 5th strand (V104-T110) of the central β-sheet of an adjacent monomer (Figure 29A and B and Figure 26B). This arrangement situated *C*-terminal residues Lys-350 and Glu-349 of one molecule near the substrate binding site of a second molecule to potentially stabilize the active site and participate in substrate binding (Figure 28B).

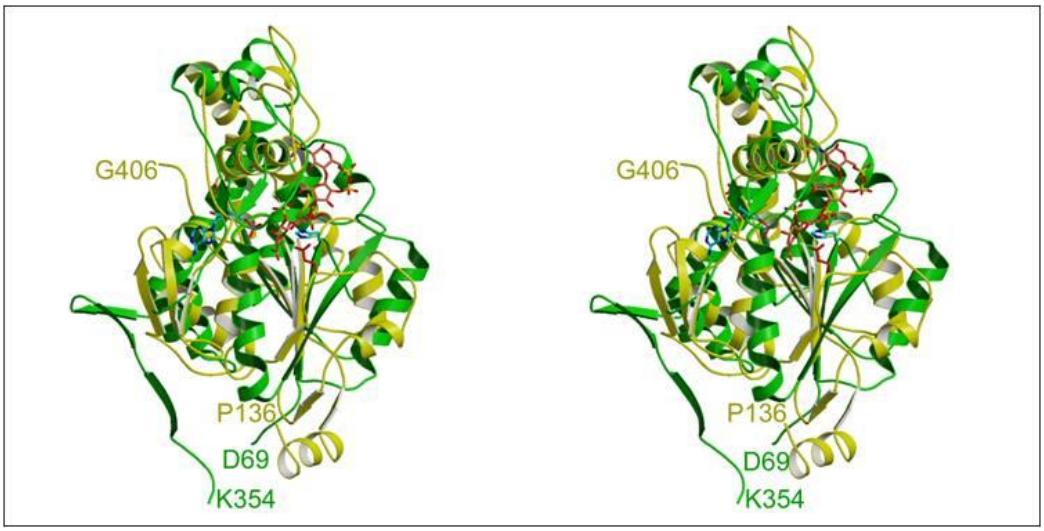


Figure 27. Stereo diagram of the superposition of 2OST and 3OST-3. Superposition of 2OST (green with PAP and His-142 in cyan) and 3OST-3 (yellow with catalytic Glu-184 in red) with bound heparan sulfate tetrasaccharide (coral) reveals global similarities and specific differences between the two enzymes. Most notable is the extended *C*-terminus of 2OST involved in trimer formation.

Comparison to Cytosolic Sulfotransferases

HS sulfotransferases, like 2OST, are type II transmembrane Golgi-associated enzymes however cytosolic sulfotransferases, such as estrogen sulfotransferase (EST), have a buried hydrophobic binding pocket for the acceptor substrate such as a hydrophobic steroid (235). Sequence alignments of 2OST with the cytosolic sulfotransferase EST revealed a slightly higher sequence identity of ~16.5% (80). Interestingly, 7 of the 45 sequence conserved residues came from the long α-helix (V182-L188) equivalent to helix 6 in EST, that runs across the top of both the donor and acceptor substrate binding pockets (Figure 28A). Structural comparison of 2OST to EST revealed the absence of a catalytic glutamate, which was noticeably present in NST1 and the 3OSTs. Instead, 2OST and EST utilized a histidine at the end of the 4^{th} strand of their central β-sheet (80). Like His-108 in EST, a structurally equivalent histidine (His-142) was present in 2OST (Fig 28A and B), which had been identified by Xu *et al.* (104) based on comparisons with cytosolic enzymes using sophisticated sequence alignments and mutational analysis. The His-142 residue for 2OST existed in two apparent conformations where one conformation positioned the side chain away from the substrate binding cleft (conformation A) while the other conformation (conformation B) superimposed very well with His-108 of EST. Conformation B positioned the histidine side chain in the active site cleft for a possible role in catalysis (Fig 28A).

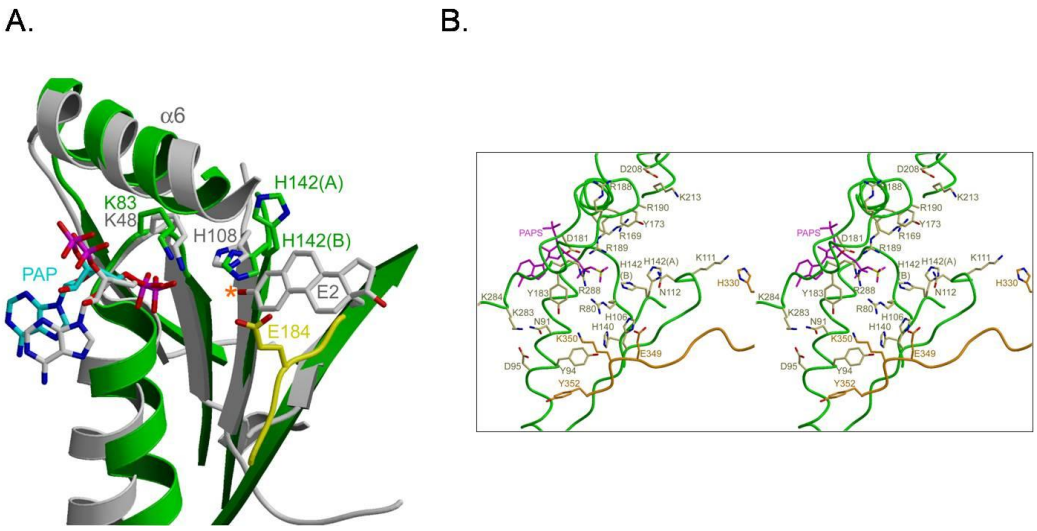


Figure 28. Active site residues of 2OST. (A) Superposition of the active site of 2OST (green) with PAP (cyan) and the cytosolic EST enzyme (gray) with PAP and acceptor substrate 17β-estradiol (E2). The figure displays the position of the acceptor OH of E2 with respect to the proposed catalytic base His-108 in EST. The structurally equivalent His-142 of 2OST is shown in the 2 apparent conformations displayed in the crystal structure. In contrast, the position of the proposed catalytic base in 3OST-3 (yellow), Glu-184, is shown originating from a different location in the structure, yet the acceptor (data not shown) location is similar to that of the OH of E2 in EST (marked by *). **(B)** Stereo diagram of the proposed acceptor binding cleft of a 2OST molecular (green) and the adjacent C-terminal tail of another 2OST molecule in the trimer (orange). PAPS (purple) is based on superposition of PAPS from the EST structure with bound PAPS (Protein Data Base ID code 1HY3) onto the PAP molecule in 2OST. Residues shown lining the cleft were mutated in this work (Table 1).

Trimeric Complex of 2OST

The most distinct structural difference between 2OST and the other HS sulfotransferases was that 2OST appeared to function as a trimer. In the crystal structure of MBP-2OST, a 3-fold crystallographic axis ran through the middle of the trimer (Fig 29A and B). Supporting this quaternary structure was the previous study in Figure 25, demonstrating that the fusion protein behaved as a trimer in solution as determined by gel filtration chromatography. The results were consistent with a study by Kobayashi *et al.*, demonstrating that 2OST isolated from Chinese hamster ovary cells migrated as an oligomer by gel filtration chromatography (99).

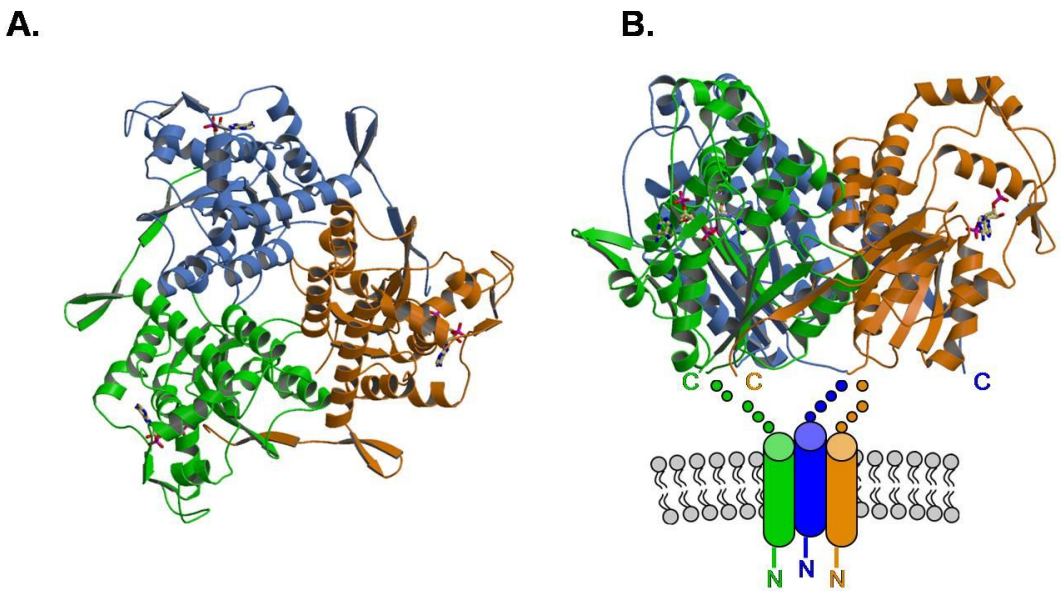


Figure 29. Trimeric complex of the catalytic domain of 2OST. (A) View of the 2OST trimer looking down the 3-fold crystallographic axis (monomers are colored green, blue, and orange). The active sites containing PAP (khaki) are located on the outer surface of the trimer. **(B)** Side view of the trimer based on a ~90º rotation of A, along an axis in the plane of the paper. The hypothetical positions of the transmembrane N-terminal helices and the Golgi membrane with respect to the trimer have been sketched into the figure.

The orientation of the 2OST monomers within the trimeric complex positioned the *N* termini of all three molecules on one side of the trimer in close proximity to one another (Fig 29B). This orientation was consistent with a type 2 membrane-bound protein where all three molecules would be anchored to the membrane via their *N*-terminal transmembrane domains (L12-L27) although this region was not present in the crystal structure. Positioning of the monomers within the trimer separated the active sites from one another, suggesting that the individual active sites may function independently (Fig 29A).

Previous reports indicated that C_5-epi formed a complex with 2OST *in vivo* (96) however it is unlikely that C_5-epi could occupy one of the monomeric sites of the 2OST trimer. In fact, the 2OST originally purified from Chinese hamster ovary cells, although

trimeric, did not contain C_5-epi (99). The proposed *in vivo* complex could involve a supramolecular structure between the trimeric 2OST and C_5-epi.

Structurally Guided Mutagenesis Study of 2OST

Site-directed mutagenesis studies were performed to identify the roles of specific amino acid residues in the catalytic function and substrate specificity of 2OST (Table 1). Amino acid residues that are involved in binding to PAP/PAPS were not the focus of this study because they were reported in a previous publication (104). Instead, amino acid residues lining the proposed substrate binding cleft were chosen for mutational analysis. Mutant enzymes were exposed to two polysaccharide substrates differing in the structure of the disaccharide repeating unit. Completely de-*O*-sulfated *N*-sulfated (CDSNS) heparin predominantly has the repeating disaccharide of -IdoUA-GlcNS-, while *N*-sulfo heparosan consists of a repeating structure of -GlcUA-GlcNS- (Fig 30). Testing the activity of the 2OST mutants with these two different substrates permitted us to identify residues that differentiate between IdoUA and GlcUA acceptor substrates in the polysaccharide chain.

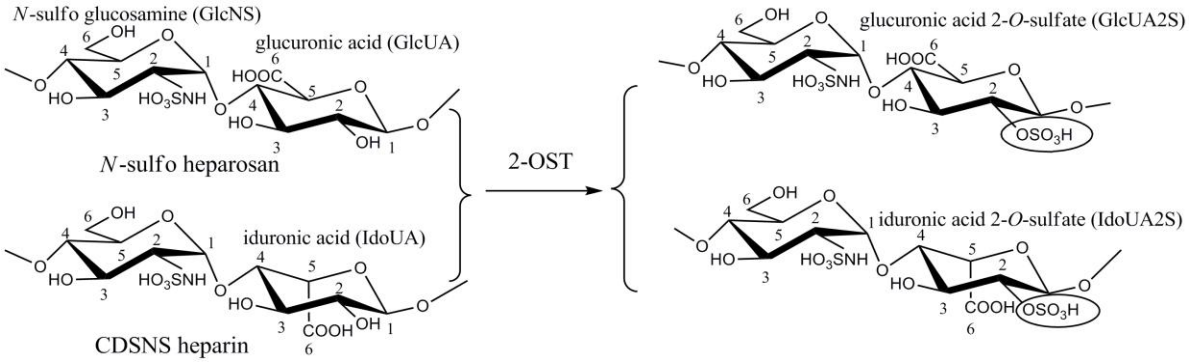


Figure 30. Reactions catalyzed by 2OST. 2OST sulfates either the GlcUA unit present in *N*-sulfo heparosan or the IdoUA unit present in CDSNS-heparin. If the polysaccharide contains both GlcUA and IdoUA units, 2OST prefers IdoUA.

Table 1. Activity of 2OST WT and mutants toward polysaccharide substrates

2OST Mutants	Sulfotransferase Activity to Polysaccharide Substrates		2OST Mutants	Sulfotransferase Activity to Polysaccharide Substrates	
	CDSNS-Heparin (-IdoUA-GlcNS-)$_n$	*N*-Sulfo Heparosan (-GlcUA-GlcNS-)$_n$		CDSNS-Heparin (-IdoUA-GlcNS-)$_n$	*N*-Sulfo Heparosan (-GlcUA-GlcNS-)$_n$
	%	%		%	%
WT	100	100	WT	100	100
R80A	1	2	R188A	100	100
N91A	51	42	R189A	0.3	100
Y94A	90	7	R190A	100	74
D95A	80	32	D208A	76	42
H106A	51	0.4	K213A	100	56
K111A	100	62	K283A	100	100
N112A	97	54	K284A	68	42
H140A	7	< 0.1	R288A	5	10
H140N	52	14	H330A	100	100
H142A	< 0.1	< 0.1	E349A	31	77
H142N	20	3	K350A	22	7
R169A	1	0.3	Y352A	69	43
Y173A	80	50	V332STP	2	0.5
D181A	20	11	L342STP	11	3
Y183A	8	2			

*The 2OST mutants were prepared using a modified site-directed mutagenesis protocol from Stratagene. The protein expression level of the mutant proteins is comparable with the wild type protein as determined by the intensity of the Coomassie blue-stained protein band that migrated at 70 kDa on SDS-PAGE.

†The activity of 2OST was assayed by incubating the purified mutant proteins with either completely de-*O*-sulfated *N*-sulfated heparin (CDSNS-heparin) [(-IdoUA-GlcNS-)$_n$] or *N*-sulfo heparosan [(-GlcUA-GlcNS-)$_n$] and [^{35}S]PAPS and the resultant ^{35}S-substrate was quantified by DEAE chromatography where 100% activity represents the transfer of 155 pmol of sulfate per µg of protein for CDSNS-heparin and 65 pmol of sulfate per µg of protein for *N*-sulfo heparosan under the standard assay conditions.

Residues Involved in Enzyme Catalysis

Based on the crystal structure, four residues (Arg-80, His-140, His-142, Arg-288) were located near the point of sulfo transfer and thus may be involved in catalysis and/or acceptor substrate binding. Consistent with previous alanine scanning mutagenesis experiments, our results demonstrated that these residues were important for activity (104). In the crystal structure, Arg-80 and Arg-288 were located near the proposed position of the sulfo moiety on the PAPS molecule and therefore could be involved in stabilization of the transition state and/or substrate binding (Fig 28B). The R80A and R288A mutants displayed

10% or less activity toward both polysaccharide substrates (Table 1). Preliminary mutagenesis studies using CDSNS-heparin demonstrated similar reduction in sulfotransferase activity while maintaining binding affinity to PAP, suggesting a potential role for these residues in substrate binding or catalysis (104).

Previously, it was demonstrated that the mutations H140A and H142A greatly reduced the activity of 2OST and that the double mutant H140A/H142A resulted in complete loss of activity (104). The crystal structure revealed that His-142 is in position to potentially serve as a catalytic base facilitating the deprotonation of the acceptor substrate, whereas His-140 was located more distal to the point of sulfo transfer and was in position to form a hydrogen bond with Arg-80. The H142A mutant of chicken 2OST had no detectable activity with the two substrates, supporting its essential role as a catalytic base, while 2OST H140A retained about 7% of wild type activity against the IdoUA containing substrate (CDSNS-heparin). To further probe the functions of His-140 and His-142, mutants H140N and H142N were prepared. The H140N mutant maintained 52% activity against the IdoUA containing substrate and 14% against the GlcUA containing substrate (*N*-sulfo heparosan). Based on the crystal structure, it appeared that the asparagine substitution could still form a hydrogen bond with Arg-80. Thus, the major role of His-140 may be to properly position Arg-80. Surprisingly, the H142N mutant maintains 20% and 3% of wild type activity against IdoUA and GlcUA containing substrates, respectively. In contrast, the equivalent mutation of His-107 of human EST to asparagine completely abolished the sulfotransferase activity (80). One possible explanation for the residual activity of the H142N mutant versus H142A could be that a water molecule occupies the location of the Nε2 atom of the histidine side chain, stabilized by interactions with Asn-142, and substituted as a weak proton acceptor in

the absence of the histidine side chain. An additional possibility was that the reaction for 2OST is more dissociative in nature, compared to EST, with greater emphasis on bond breaking and stabilization of the leaving group rather than deprotonation of the nucleophile, decreasing the need for a strong base.

Residues Involved in Trimer Formation

Several residues along the *C*-terminal tail were subjected to mutational analysis including His-330, Glu-349, Lys-350, and Tyr-352 (Fig 28B). These residues were selected as they from hydrogen bonding interactions with the β-strand of the adjacent monomer. The only residue showing a loss of activity greater than 70% was the Lys-350 mutant (Table 1). The side chain of this residue was extended into the active site of the adjacent molecule, suggesting it could play a minor role in substrate binding.

Although no single point mutation along the *C*-terminal tail caused a severe decrease in activity, introduction of a premature stop codon after either Val-332 or Leu-342 resulted in a drastic loss in activity toward both substrates. These data suggested that deletion of residues along the *C*-terminus eliminated the intermolecular interactions between the protein backbone atoms, disrupting the adjacent active site and potentially caused destabilization of the protein and/or the protein trimeric complex. To determine if the truncation mutant disrupted the trimeric complex, the V332STP mutant protein was analyzed by gel filtration chromatography. The chromatogram for the wild type MBP-2OST revealed three distinct peaks (Fig 31A) representing aggregated MBP-2OST (peak 1: fractions 3-9) in the void-volume, trimeric MBP-2OST (peak 2: fractions 12-16) (sample used for crystallization trials), and MBP (peak 3: fractions 23-28) respectively. Of these peaks, the only containing sulfotransferase activity was that consistent with the trimer.

The mutant (2OST V332STP) revealed a starkly different elution profile from the wild type MBP-2OST construct. The aggregated protein eluted with similar amount and position as the normal wild type MBP-2OST, however there was a substantial decrease in the intensity of the trimeric 2OST peak and the appearance of an additional peak before MBP between fractions 19 and 23 (Fig 31B). This peak corresponded to a molecular weight of approximately 70 kDa and represented monomeric MBP-2OST (Fig 31C). Sulfotransferase activity was detected predominately within the trimeric peak as well as the monomeric 2OST peak, but at a lower level. The activity for the mutant trimer was substantially lower than for the wild type protein, suggesting the extended *C*-terminal tail influenced trimer formation and was crucial for the activity.

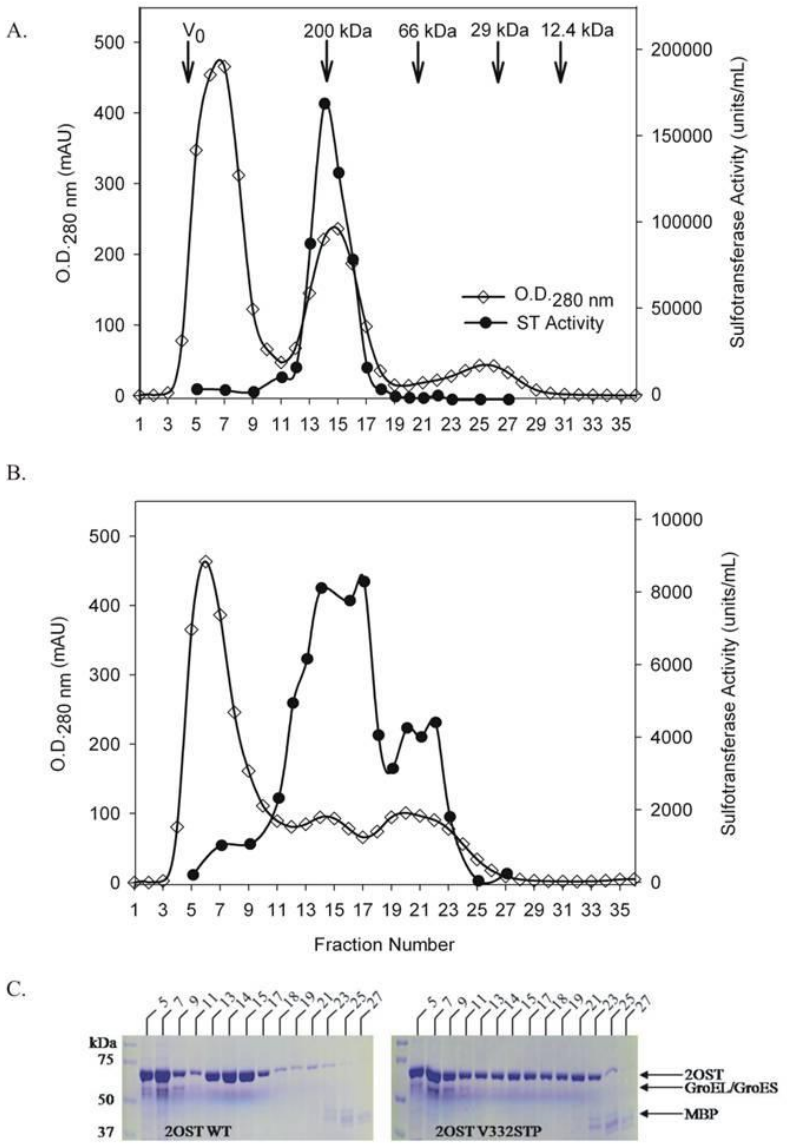


Figure 31. Comparison of the oligomeric state of 2OST WT to V332STP. Approximately 7 mg amylose-agarose column purified wild type or premature stop codon mutant (V332STP) 2OST were analyzed by gel permeation chromatography. The sulfotransferase activity was also determined using CDSNS Heparin as a substrate. Both the UV absorbance and sulfotransferase activity are plotted against fraction number for wild-type **(A)** and mutant **(B)** 2OST. The molecular weight standards resolved under the same experimental conditions are indicated by *arrows*. One unit is equivalent to one picomole of sulfate transferred to substrate under standard reaction conditions. Panel **C** shows specific fractions that were analyzed by SDS-PAGE using a 10% Tris-HCl gel stained with Coomassie Blue for both the 2OST WT and V332STP. The fractions analyzed are labeled with a number above each lane of the SDS gel. SDS-PAGE analysis of these fractions from gel permeation chromatography confirmed the presence of 2OST within all fractions from 5 to 23.

Residues Involved in Substrate Binding and Specificity

Residues Tyr-183, Arg-169, Asp-181, and Arg-189 were located in the 2OST substrate binding cleft but distal to the catalytic core implicating them in acceptor substrate binding (Fig 28B and Fig 32). Mutations of these residues showed a noticeable effect on enzymatic activity (Table 1). Of these mutations, Y183A and R169A resulted in a substantial loss of activity toward both CDSNS heparin and *N*-sulfo heparosan. Tyr-183 was located toward the predicted non-reducing end of the binding pocket based on comparisons to the crystal structure of 3-OST-3 (63). The side chain of Tyr-183 extended into the solvent, suggesting it does not play a structural role but rather may play a significant role in substrate binding. The side chain of Arg-169 was located on the reducing end of the binding pocket. Most of this side chain was buried with the guanidinium moiety on the surface of the binding pocket. Arg-169 is conserved in NST-1 and the 3-OST isoforms as well. In the 3-OST-3 structure with a tetrasaccharide bound, this residue (Arg-248) did not directly interact with the substrate (63). The main role of Arg-169 may be to stabilize the PSB-loop as the guanidinium moiety is within hydrogen bonding distance to the backbone carbonyl oxygen of Pro-82 in this loop. Also lining the pocket, but largely buried, is the side chain of Asp-181. This residue is in position to form hydrogen bonds with Arg-189 (2.8 Å) and/or Arg-288 (2.9 Å). Such interactions suggested that Asp-181 may help position Arg-288 and Arg-189 for substrate binding and/or catalysis. As with Arg-288, replacing Asp-181 with alanine resulted in a substantial loss of the sulfotransferase activity toward both CDSNS heparin and *N*-sulfo heparosan (Table 1).

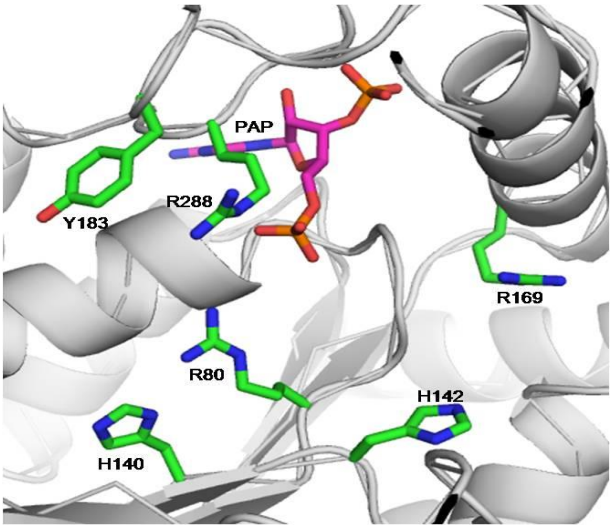


Figure 32. Amino acid residues involved in substrate binding/catalysis. A closer view of the active site of 2OST is shown with the secondary structure of 2OST (gray) bound to PAP (magenta) bound. The residues involved in substrate binding/catalysis including Tyr-183, Arg-288, His-140, Arg-80, His-142, and Arg-169 are colored green.

Residues Involved in Redirecting Substrate Specificity

Although wild type 2OST sulfates both GlcUA and IdoUA in HS substrates, the mutants Y94A, H106A and R189A showed substantial preference for one substrate over the other. Both Y94A and H106A mutants sulfated CDSNS heparin, the IdoUA containing substrate, but lost activity with *N*-sulfo heparosan, the GlcUA containing substrate, displaying 10 fold and 100 fold differences in percent activity, respectively. These results suggest that Tyr-94 and His-106 allow recognition of the GlcUA containing substrate. Thus, replacing either residue with alanine resulted in a substantial preference for the polysaccharide substrate containing IdoUA.

Most impressive was the selective preference of R189A for the GlcUA containing substrate over IdoUA. The mutant R189A retained full activity with the GlcUA containing substrate but showed virtually no activity with the IdoUA containing substrate. This observation suggested that residue Arg-189 is likely important for recognition of IdoUA

containing substrate and replacing the arginine residue with an alanine does not allow the enzyme to exhibit substrate flexibility.

To further examine the role of Arg-189 in substrate recognition, we reconstituted the biosynthesis of IdoUA2S *in vitro*. An unepimerized polysaccharide substrate sulfated only on the *N*-position of the glucosamine unit (similar to *N*-sulfo heparosan (18)), was incubated with wild type 2OST and 2OST R189A in the presence or absence of C_5-epi, where C_5-epi converted part of the GlcUA to IdoUA (as illustrated in Fig 33C). The resultant polysaccharides were then characterized by disaccharide analysis of the products (Fig 33A, 33B). As expected, the product modified by wild type 2OST predominantly contained GlcUA2S in the absence of C_5-epi (Fig 33A, red line) or IdoUA2S in the presence of C_5-epi (Fig 33A, blue line) (98,236). In contrast, the products modified by 2OST R189A contained only GlcUA2S regardless of the absence or presence of C_5-epi (Fig 33B), suggesting that the mutant protein did not recognize IdoUA as an acceptor unit in the polysaccharide substrate.

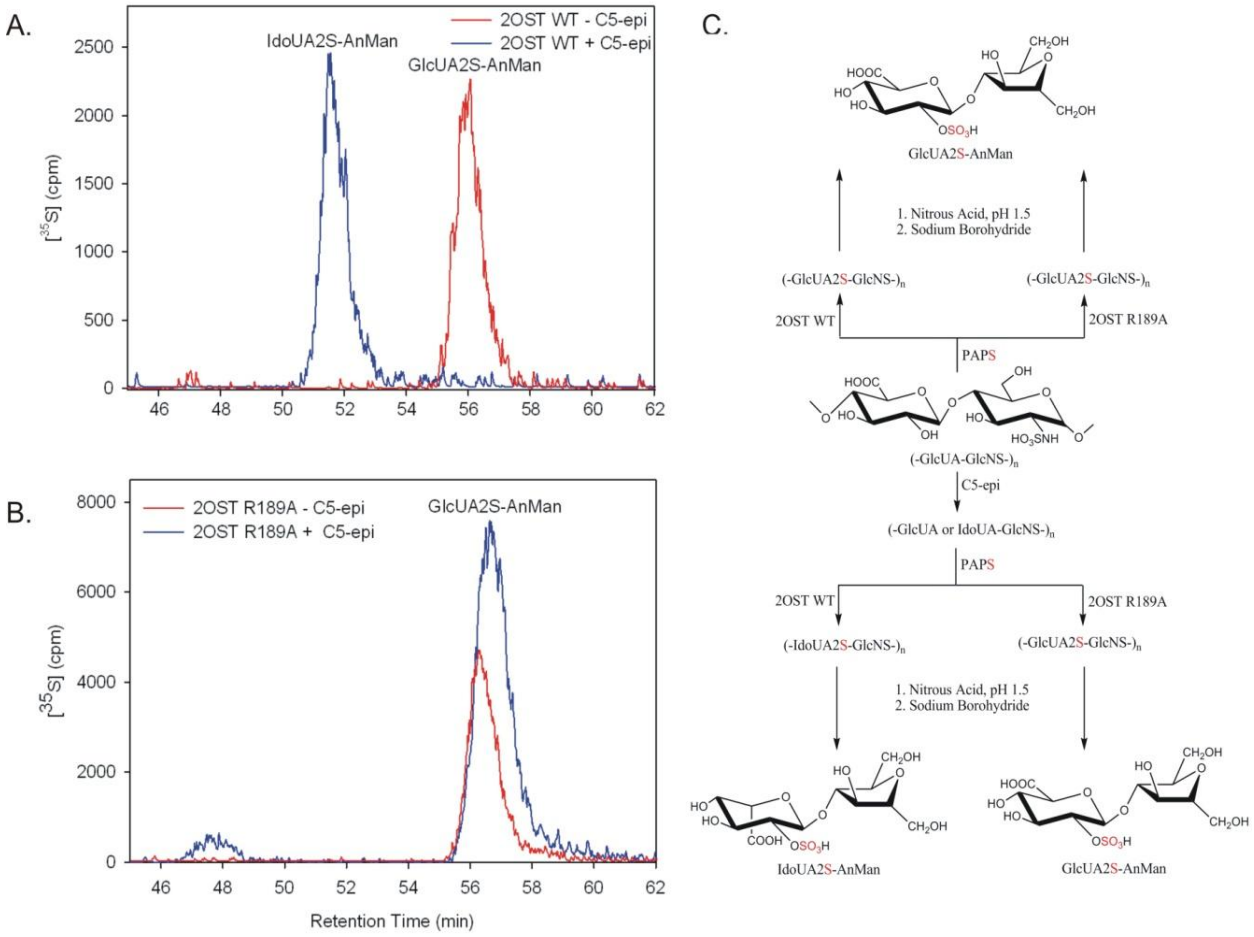


Figure 33. RPIP-HPLC chromatograms of the disaccharide analysis of 2OST wild type and 2OST R189A-modified polysaccharides. (A) Chromatogram of nitrous acid degraded 2-O-[^{35}S]sulfated *N*-sulfo heparosan using 2OST WT enzyme in the presence (*blue line*) or absence (*red line*) of C_5-epi modification. **(B)** Chromatogram of nitrous acid degraded 2-O-[^{35}S]sulfated *N*-sulfo heparosan using 2OST R189A enzyme with (*blue line*) or without (*red line*) C_5-epi modification. **(C)** Reactions involved in preparing 2OST-modified polysaccharide with or without C_5-epi and degradation by nitrous acid. AnMan represents 2,5-anhydromannitol.

The roles of Tyr-94, His-106, and Arg-189 in directing the substrate specificity are currently unknown since a crystal structure of a ternary complex containing an appropriate oligosaccharide is currently unavailable. All three mutants of Y94A, H106A, and R189A were analyzed by gel filtration chromatography alongside the wild-type to identify any key differences in the overall conformation of the protein. The 2OST wild type and mutant proteins presented elution profiles consistent with the MBP-2OST protein used during crystallization (Fig 34). Several fractions were analyzed for sulfotransferase activity, revealing activity solely within the trimeric 2OST peak for wild-type as well as all mutant 2OSTs. This additional information suggested that the mutant enzymes did not cause changes in the overall conformation of the enzyme since all mutants including the wild-type 2OST maintained the highly active trimeric structure.

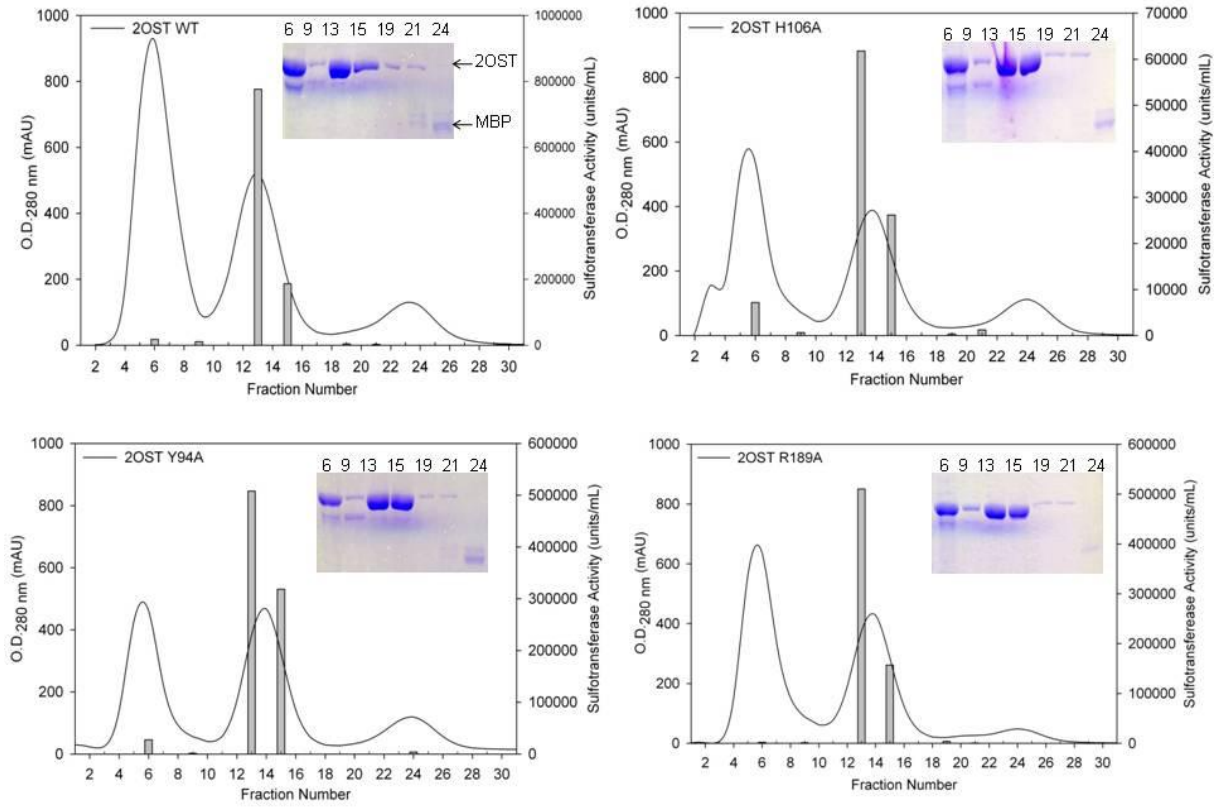


Figure 34. Comparison of the oligomeric state of 2OST WT with 2OST Y94A, H106A, and R189A. Approximately 20 mg of amylose-agarose column purified WT or mutant (Y94A, H106A, and R189A) 2OST was analyzed by gel permeation chromatography. The sulfotransferase activity was determined by using CDSNS heparin as a substrate, and 2OST R189A was analyzed using *N*-sulfo heparosan as a substrate. The UV absorbance (solid line) and sulfotransferase activity from selected fractions (bar graph) are plotted against fraction numbers for 2OST WT and mutants. One unit is equivalent to 1 pmol of sulfate transferred to the substrate under standard reaction conditions. Some fractions possessed sulfotransferase activity <100 units/mL and thus cannot be seen in the bar graph because of axis scaling. (Insets) Fractions analyzed for sulfotransferase activity were also analyzed by SDS-PAGE, with a number above each lane of the picture of the SDS gel representing fractions from the gel filtration column.

Conclusions

The biosynthesis of HS involves a series of specialized enzymes, including glycosyl transferases, an epimerase, and various sulfotransferases. Modification by 2-*O*-sulfotransferase, which transfers a sulfo group to both GlcUA and IdoUA units with a strong preference for IdoUA, has been implicated in a variety of biological functions (65). The functions of 2-*O*-sulfated iduronic acid (IdoUA2S) in HS have been demonstrated by *in vitro*

and *in vivo* studies. Unlike IdoUA2S, GlcUA2S is much less abundant in the HS isolated from natural sources and thus the function of GlcUA2S is currently unknown. The IdoUA2S unit is necessary for binding to numerous growth factors and is crucial for triggering fibroblast growth factor-mediated signal transduction pathways (105). The sequence of 2OST is highly conserved across most vertebrate genomes including human and chicken, suggesting a central physiological role in these organisms. Several animal knockout studies have demonstrated an essential role for this enzyme in animal development (101,103,124). With that said, it is necessary to gain a more detailed understanding of the structure-function relationship of 2OST in order gain more insight into its role in animal development.

Here, the crystal structure of 2OST was reported as a MBP fusion protein in complex with PAP. Unlike previously characterized HS sulfotransferases, 2OST is present in a trimeric form that appears to be essential for enzymatic activity. Using the crystal structure as a guide, several amino acid residues along the active site that may play a potential role in substrate recognition were mutated. These mutants were analyzed for sulfotransferase activity against two polysaccharide substrates differing in repeating disaccharide units and compared to the wild type 2OST enzyme. Several residues were identified that may be essential for substrate binding/catalysis including Arg-80, His-140, His-142, Arg-169, Tyr-184, and Arg-288 with mutation resulting in 10% or less activity towards both substrates compared to the wild type.

One unique feature of 2OST compared to the other HS sulfotransferases is the presence of an extended *C*-terminal tail that forms an anti-parallel strand arrangement with the 5^{th} strand of the central β-sheet of an adjacent monomer. Mutation of several residues including His-330, Glu-349, Lys-350, and Tyr-352 did not demonstrate a significant loss in

activity compared to the wild type 2OST. However, introduction of a premature stop codon after either Val-332 or Leu-342, eliminating the *C*-terminal tail residues, resulted in a dramatic loss in sulfotransferase activity. This result suggests that removal of the *C*-terminal tail disrupts the adjacent active site and potentially destabilizes the trimeric complex of 2OST. To further explore this theory, the 2OST V332STP mutant was analyzed by gel filtration chromatography and compared to the wild type enzyme. The V332STP mutant revealed a significant reduction in the MBP-2OST trimer and the appearance of a monomeric MBP-2OST peak, suggesting a disruption in the trimeric complex. Analysis of sulfotransferase activity of these two peaks detected activity predominately within the trimeric peak. The activity for the trimer within the V332STP mutant was significantly lower compared to the wild type (8000 U/mL versus 160,000 U/mL), suggesting that the *C*-terminal tail is essential for the formation of the trimer and sulfotransferase activity.

Through our mutational analysis, three mutants of Y94A, H106A, and R189A were identified that demonstrated a selective preference for one substrate over another. Both Y94A and H106A selectively sulfated the IdoUA-containing substrate (CDSNS-heparin) whereas R189A selectively sulfated the GlcUA-containing substrate. To ensure that these three mutants were not causing global changes of the 2OST protein, all three mutants alongside the wild type were analyzed by gel filtration chromatography, determining that all the mutants migrated as a trimer which possessed all sulfotransferase activity. These results suggest that the mutations do not lead to large scale structural changes that disrupt the active trimer, supporting a local change within the active site that causes the differences in substrate specificity.

Demonstrating over 300 fold difference in percent activity, the role of R189A mutant was further explored by reconstitution of the biosynthesis of IdoUA2S *in vitro* through modification of *N*-sulfo heparosan with 2OST R189A in the absence or presence of C_5-epi followed by disaccharide analysis using nitrous acid degradation at pH 1.5. This process was repeated using wild type 2OST for comparison. Products modified by the wild type 2OST predominately contained GlcUA2S-AnMan disaccharides in the absence of C_5-epi but solely IdoUA2S-AnMan disaccharide products in the presence of C_5-epi. The products modified by the R189A mutant, on the other hand, contained only GlcUA2S-AnMan disaccharide products regardless of the presence or absence of C_5-epi, confirming the mutagenesis results. These results suggest that 2OST R189A solely recognizes GlcUA units in the presence of IdoUA, which is less likely with wild type 2OST. One potential application of this mutant would be to synthesize therapeutic heparin with reduced side effects. Heparin-induced thrombocytopenia is a major side effect of heparin, caused in part by the binding to platelet factor 4. It is known that the IdoUA2S unit, the product of wild type 2OST modification, is involved in heparin binding to platelet factor 4 (237). Using the 2OST R189A mutant to synthesize heparin, the IdoUA2S unit could be replaced with GlcUA2S, hopefully reducing the side effect. Another potential application of the engineered mutant would be to produce HS and heparin with highly enriched GlcUA2S units for probing the biological function in glycomics studies (238).

In conclusion, our results have provided structural and biochemical information for delineating the mechanism of action of 2OST and evidence for altering the substrate specificity of 2OST for the synthesis of HS with unique biological functions.

Chapter IV

Determining the Activity of Heparan Sulfate C_5-Epimerase Using Engineered 2OST

Heparan sulfate consists of glucuronic acid (or iduronic acid) linked to glucosamine carrying various sulfo groups. The HS C_5-epimerase (C_5-epi) is responsible for conversion of glucuronic acid to iduronic acid. The method for determining the C_5-epi activity has been cumbersome due to the use of a site-specifically ^{3}H-labeled polysaccharide substrate. Here, a two-enzyme coupling assay was reported to determine the activity of C_5-epi. Using the engineered 2OST mutants identified in Chapter III, an effective assay was designed to monitor the activity of heparan sulfate C_5-epi (239). Unlike wild-type 2OST which recognizes both glucuronic acid and iduronic acid, the 2OST Y94I mutant transfers sulfate to the iduronic acid residue but not glucuronic acid. The substrate selectivity allowed determination of the amount of epimerase activity by simply monitoring the incorporation of 2-*O*-sulfo groups onto the IdoUA units by 2OST Y94I. Our approach will significantly reduce the complexity for assaying the activity of C_5-epi and facilitate the structural and functional analysis of C_5-epi.

Redirecting the Substrate Specificity of 2OST

Upon the discovery of the engineered 2OST mutants, there was an opportunity to take advantage of their selective properties for the development of an effective assay for C_5-epi activity. The current method for determining the activity for C_5-epi is less straightforward due to the use of a site-specifically tritium labeled polysaccharide substrate. Campbell *et al.* reported a method to measure the epimerase activity by determining the released tritium that was introduced at the C_5 position of the GlcUA/IdoUA unit of CDSNS-heparin (93). The C_5-epi de[^{3}H]protonates at the C_5-position of a GlcUA or IdoUA unit, dissociating the [^{3}H]-label from the polysaccharide. Although this method was sufficient to guide the purification of C_5-epi from bovine liver (240), a large amount of [^{3}H]H_2O must be used to prepare the tritium labeled polysaccharide which cannot be readily completed in a standard academic laboratory. Another issue with the current assay is that the reversibility of the enzyme causes continuous release and addition of the ^{3}H-label therefore the measurement could be incorrect. The need for a ^{3}H-labeled polysaccharide substrate was bypassed to assay the C_5-epi activity by using a 2OST mutant with altered substrate specificity (Fig 35C). The substrate used in this method was *N*-sulfo heparosan, a polysaccharide containing solely GlcUA units. The 2OST Y94A mutant was engineered to specifically sulfate IdoUA therefore the 2OST mutant will not sulfate *N*-sulfo heparosan unless the GlcUA units have been converted to IdoUA by C_5-epi.

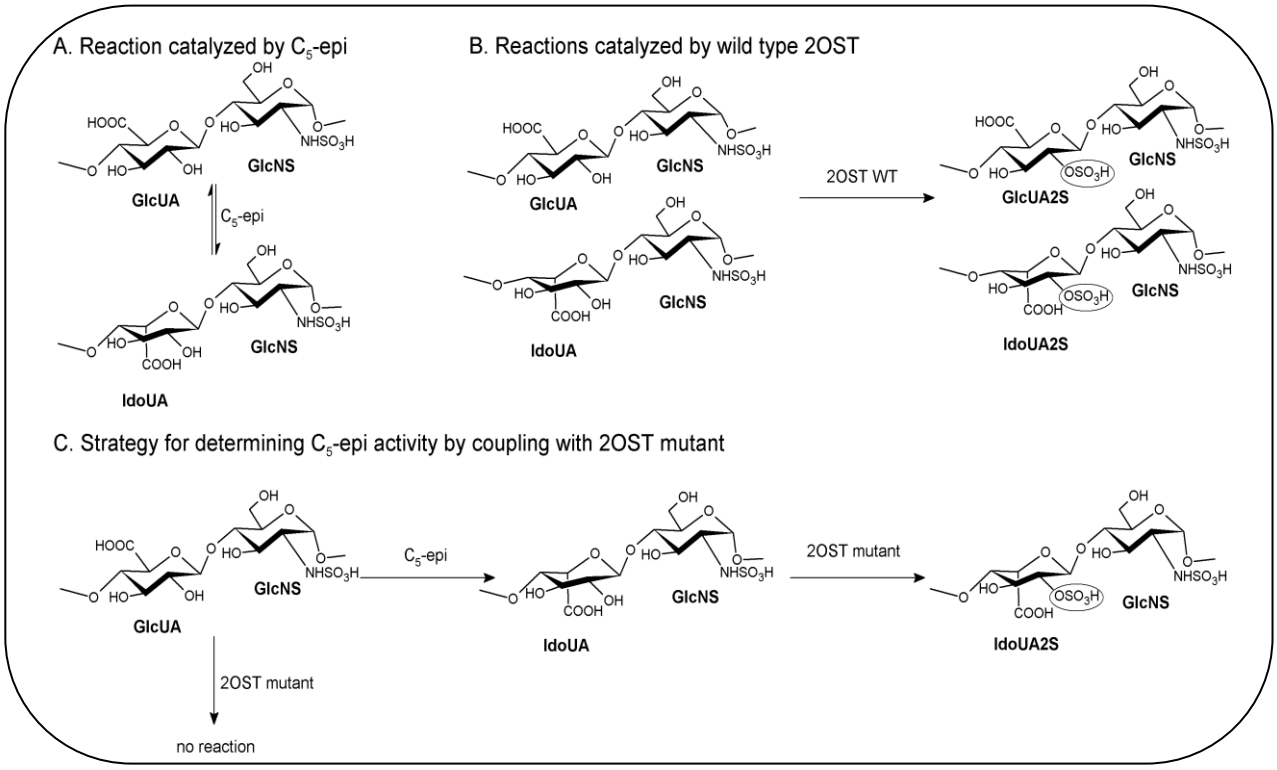


Figure 35. Chemical reactions catalyzed by C_5-epi and 2OST. (A) HS C_5-epi catalyzed reaction. HS C_5-epi catalyzes the conversion of GlcUA unit in HS to IdoUA and the reverse reaction, namely from the IdoUA to GlcUA. **(B)** 2OST WT catalyzed reaction. 2OST WT catalyzes the 2-*O*-sulfation of both GlcUA and IdoUA units in the polysaccharide. **(C)** Strategy for determining C_5-epi activity by coupling C_5-epi and 2OST mutant. By coupling these two enzymes together, the IdoUA produced by epimerase catalyzed reaction is converted to IdoUA2S, which can be detected by the level of [^{35}S]sulfate incorporation to the polysaccharide.

The heparan sulfate C_5-epi converts GlcUA to IdoUA by abstracting the proton of the C_5-hydrogen of GlcUA followed by reinsertion of the proton (241), however the reversibility of the reaction results in the polysaccharide containing both GlcUA and IdoUA (Fig 35A). The resultant polysaccharide could be modified by wild type 2OST to generate IdoUA2S and GlcUA2S units within the chain (Fig 35B). Previously, introducing a single point mutation to the active site of 2OST altered the substrate specificity of the enzyme (59). The 2OST mutants, Y94A and H106A, selectively sulfate the IdoUA unit whereas R189A selectively sulfates the GlcUA unit. To further confirm this finding, the susceptibility of 2OST WT and

2OST Y94A was compared to *N*-sulfo heparosan, the GlcUA-containing polysaccharide across a range of 2OST concentrations. Indeed, 2OST Y94A displayed merely basal level of sulfation whereas 2OST WT demonstrated excellent activity toward *N*-sulfo heparosan (Fig 35A).

In a separate experiment using CDSNS heparin as a substrate, a polysaccharide predominately containing the IdoUA unit, both wild type and mutant 2OST exhibited similar susceptibility especially at higher concentrations (Fig 36B). Taken together, our data suggest that the 2OST mutant could be used to detect the presence of IdoUA in the polysaccharide, leading to the possibility of assaying the activity of C_5-epi as depicted in Figure 36C. The coupling of C_5-epi and 2OST is an effective approach to monitor the level of epimerization since previous studies reported that 2-*O*-sulfation prevented "back epimerization" (94,241). In other words, the addition of 2-*O*-sulfo groups will lock the configuration of the iduronic acid into place so one can effectively detect the level of conversion.

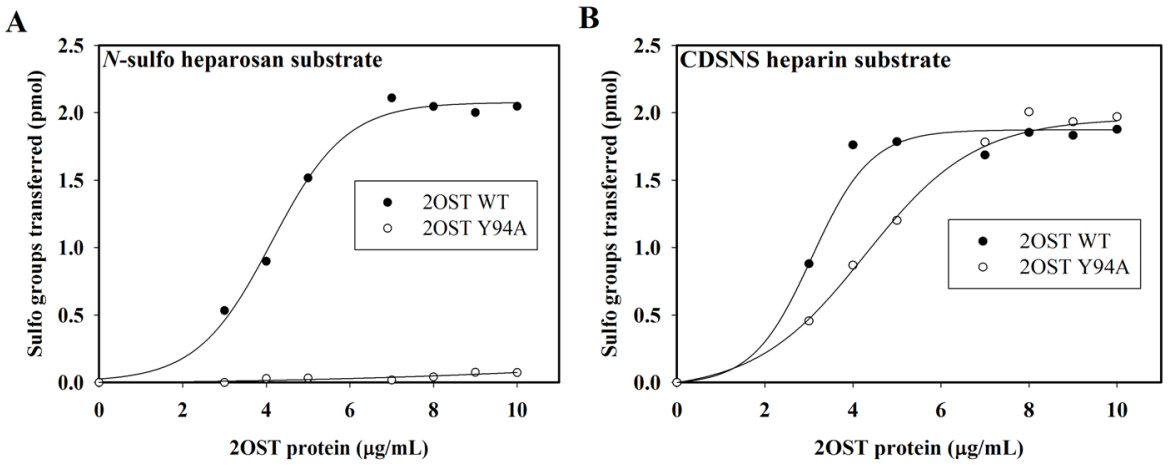
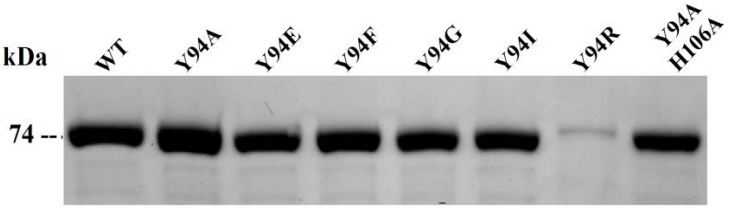


Figure 36. The sulfotransferase activity of 2OST WT and Y94A mutant toward different polysaccharides. Different amounts of 2OST proteins (0-1.0 μg) were incubated with either 2.0 μg *N*-sulfo heparosan **(A)** or 2.0 μg CDSNS heparin **(B)** and 9.9 × 10^5 cpm of [^{35}S]PAPS for 5 minutes at 37 ^{o}C. The sulfated polysaccharide was purified by DEAE resin and was determined by scintillation counting. The results for 2OST WT were shown in filled circles (-●-) and 2OST Y94A in empty circles (-o-).

Optimization of the 2OST Tyr-94 Mutant

A series of single point mutations were prepared to further optimize the selective of the 2OST Tyr-94 mutant. The Tyr-94 was replaced with the acidic residue such as glutamic acid and a basic residue such as arginine. The size of the side chain was varied by replacing with either phenylalanine or glycine. In addition, a double mutant Y94A/H106A was prepared since 2OST H106A demonstrated sulfation of IdoUA units only. All of the single point mutants were expressed on a small scale in Origami B (DE3) cells as MBP fusion proteins and purified by an amylose column as described in the Mutational Analysis section. With the exception of 2OST Y94R, the purified 2OST Tyr-94 mutants were successfully expressed as determined by SDS-PAGE analysis revealing a band at 74 kDa (Fig 37A). The 2OST mutants were analyzed for sulfotransferase activity against the two polysaccharide substrates of *N*-sulfo heparosan and CDSNS-heparin and compared to the wild type enzyme (Fig 37B).

A



B

2OST mutants	*N*-sulfo heparosan (-GlcUA-GlcNS-)$_n$ (%)	CDSNS-heparin (-IdoUA-GlcNS-)$_n$ (%)
WT	100	100
Y94A	5.3 ± 0.5	80 ± 1
Y94E	2.4 ± 0.7	19 ± 1
Y94F	2.5 ± 0.4	30 ± 4
Y94G	2.4 ± 0.1	33 ± 6
Y94I	5.1 ± 0.9	87 ± 3
H106A	9.6 ± 0.4	63 ± 6
Y94A/H106A	0.3 ± 0.2	3.1 ± 0.1

Figure 37. The expression and the sulfotransferase activity of 2OST mutants. The 2OST mutants were overexpressed in Origami-B cells using pMalc2x vector. The purified proteins were analyzed by 10% SDS-PAGE with 2OST WT **(A)**. The sulfotransferase activities of these mutants were tested using either *N*-sulfo heparosan or CDSNS heparin as a substrate and shown in percentage relative to wild type enzyme **(B)**.

Among the mutants analyzed, 2OST Y94A retained most of the sulfotransferase activity to CDSNS-heparin (80%) while retaining only 5% activity to *N*-sulfo heparosan. This result was consistent with the previous study seen in Table 1 in Chapter III. Similar results were found for the Y94I mutant, retaining activity of 5% and 87% to *N*-sulfo heparosan and CDSNS heparin, respectively. Substitution of tyrosine to phenylalanine (2OST Y94F) or glycine (2OST Y94G) also resulted in a significant loss in activity towards *N*-sulfo heparosan while maintaining substantial activity towards CDSNS-heparin, suggesting that the size of the side chain of the Tyr residue is not essential for the substrate

selectivity. Mutation to a negatively charged glutamate residue resulted in significant loss in activity toward both substrates, although its sulfotransferase activity to CDSNS-heparin decreased even further (19% of wild type 2OST). The effect of the basic residue arginine on substrate selectivity was not determined due to an insufficient quantity of enzyme for use in the activity assay (Fig 37A). The double mutant, 2OST Y94A/H106A, lost nearly all activity towards both substrates, which was highly unexpected since the single point mutants presented a preference for the IdoUA-containing substrate previously (Table 1). Since substitution of 2OST Tyr-94 with either alanine or isoleucine demonstrated similar selectivity to IdoUA units, the 2OST Y94I was utilized in the subsequent study for assaying the activity of C_5-epi.

Utilizing 2OST Y94I to Assay C_5-Epi Activity

The 2OST Y94I mutant selectively sulfates IdoUA units therefore the use of this mutant protein was employed to detect the level of IdoUA units, resulting from the action of C_5-epi (Figure 35C). The GlcUA containing substrate, *N*-sulfo heparosan, was incubated with increasing amounts of C_5-epi ranging from 0-66 µg/mL followed by the addition of either 2OST WT or 2OST Y94I in the presence of [^{35}S]PAPS. The susceptibility to 2OST Y94I modification was elevated with increasing amounts of epimerase as indicated by the amount of [^{35}S]sulfo groups incorporated into the polysaccharide substrate (Fig 38). A clear linear relationship between the amount of C_5-epi and the susceptibility to 2OST Y94I modification was observed, suggesting the validity of this method for measuring the activity of C_5-epi.

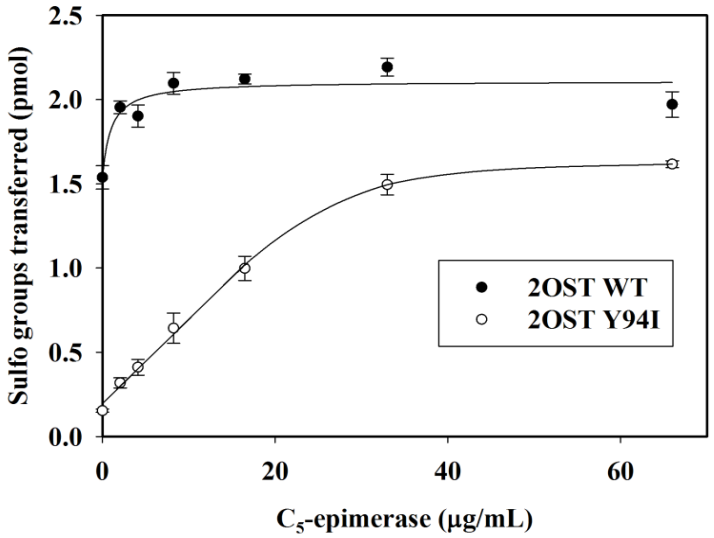


Figure 38. Dose response of C_5-epi to the [^{35}S]sulfation by 2OST Y94I and 2OST WT. Different amounts of C_5-epi (0-66 μg/mL) were incubated with 2.0 μg *N*-sulfo heparosan for 30 minutes at 37 ^{o}C. Then epimerization reactions were stopped by boiling for 5 minutes. After centrifugation, the supernatants were mixed with either 2OST WT or 2OST Y94I and [^{35}S]PAPS (4.4×10^5 cpm), followed by incubation for 5 minutes at 37 ^{o}C. The sulfated polysaccharide was purified by DEAE resin and determined by scintillation counting. The results obtained using 2OST WT were shown in filled circles (-●-). And 2OST Y94I in empty circles (-o-).

A control experiment using 2OST wild type demonstrated a high level of [^{35}S]sulfo group incorporation into *N*-sulfo heparosan even in the absence of C_5-epi, which was expected since wild type 2OST can effectively sulfate GlcUA unit (Fig 38). The wild type protein revealed a very narrow linear range of [^{35}S]sulfated product in response to epimerase, making it unsuitable for determining the activity of C_5-epi with this two enzyme-coupled method. In addition, the ability of the wild type protein to sulfate GlcUA units would make it very difficult to monitor the level of IdoUA resulting from C_5-epi using this assay since the [^{35}S]sulfo groups would be incorporated onto both GlcUA and IdoUA units.

Applying the Two Enzyme-Coupled Assay to *N*-Terminally Truncated C_5-Epi Mutants

Once the assay was established, the method was then applied to C_5-epi constructs differing in truncation sites from the *N*-terminus. As previously determined, the human C_5-epi is a 617-amino acid protein consisting of three domains, cytoplasmic (M1-K11), transmembrane (N29-N617), and catalytic domain (N29-N617) (242). The catalytic domain of C_5-epi was further probed by preparing six different *N*-terminally truncated C_5-epi constructs as MBP fusion proteins. Based on a 3D structural prediction using the PredictProtein server, each truncation site was meticulously selected from a flexible loop to avoid overall global structural changes in the protein (243). The truncated C_5-epi mutant proteins were successfully expressed as MBP fusion proteins and analyzed by SDS PAGE, revealing an expected gradual decrease in the size of the truncation mutants (Fig 39A). The C_5-epi activity with each truncation mutant was measured by the established method. The truncation mutant C_5-epi (N92-N617) demonstrated similar epimerase activity to C_5-epi (D34-N617), C_5-epi (E53-N617), and C_5-epi (Q72-N617). The further truncation mutants, C_5-epi (W203-N617) and C_5-epi (F257-N617), exhibited no epimerase activity as compared to the negative control without epimerase (Fig 39B). Taken together, these data suggested that the region of N92-W203 was important for the function of C_5-epi.

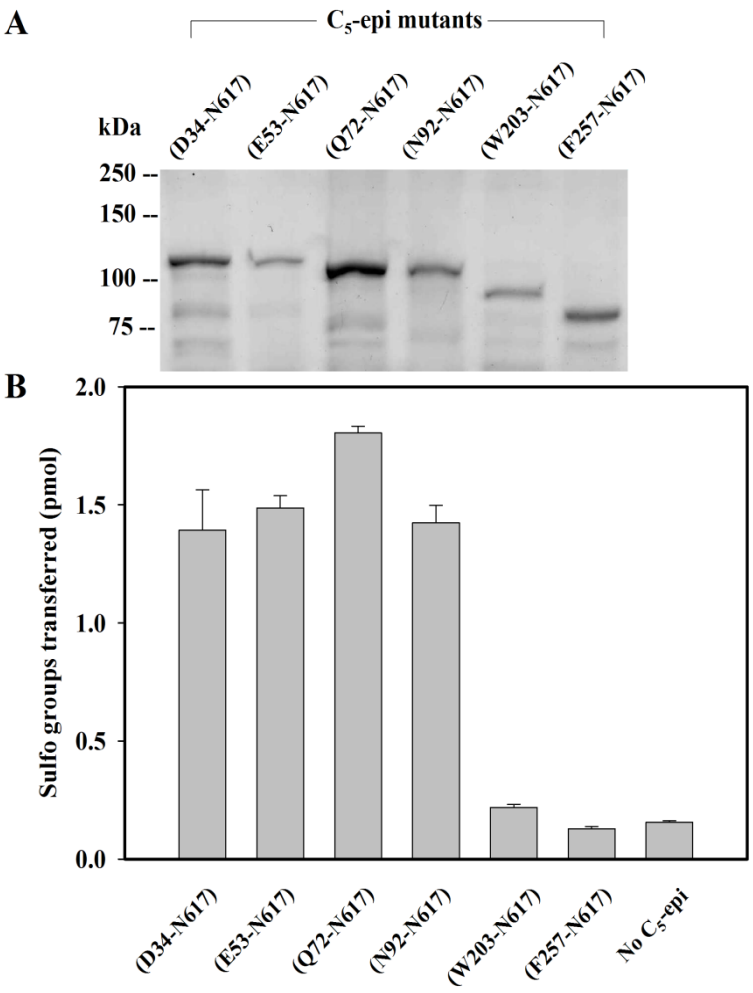


Figure 39. Activities of different truncations of C_5-epi. The different truncations of human heparan sulfate C_5-epimerase were overexpressed in Origami-B (DE3) cells using the pMalc2x vector. The picture of the purified proteins analyzed by SDS-PAGE is shown in Panel **A**. The values for the activity measured by the developed assay are shown in Panel **B**.

Understanding the Catalytic Mechanism of C_5-Epi

The proposed mechanism of C_5-epi has been compared to heparin lyases due to a similar transition state intermediate (Fig 40), a carbon anion at the C_5 position of the GlcUA/IdoUA unit, among the two enzymes (244). A tyrosine residue (Tyr-257) had been implicated in playing an essential role in the catalytic function of heparin lyases therefore it was speculated that some of the tyrosine residues could be important for C_5-epi (244).

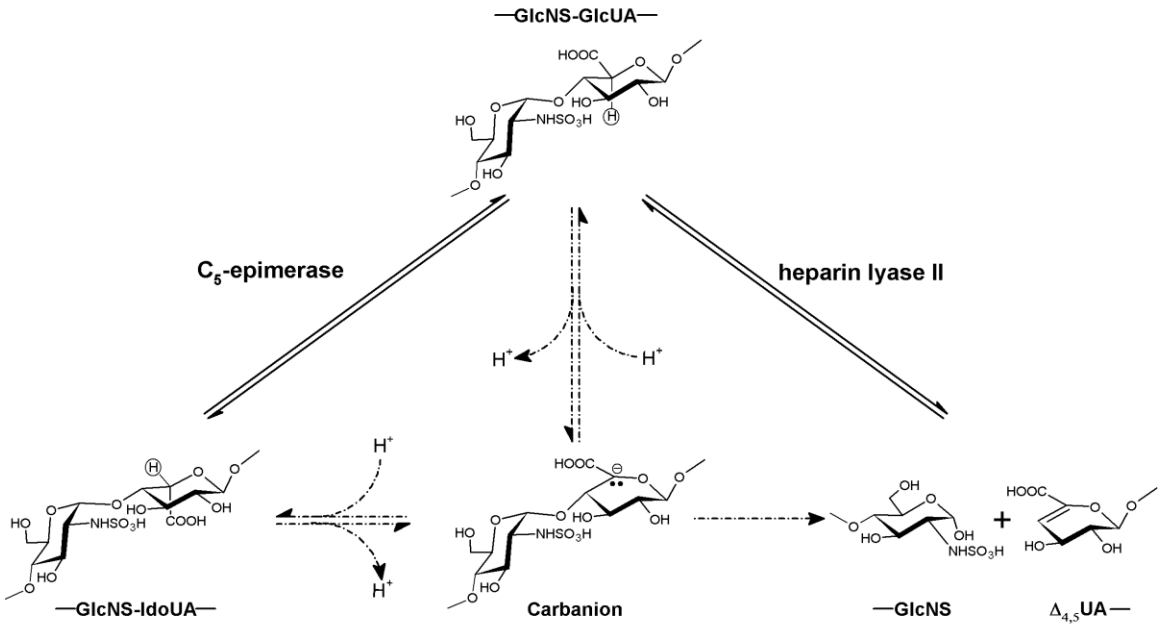


Figure 40. Proposed catalytic mechanisms of C_5-epi and heparin lyase II (244). Both C_5-epi (left) and heparin lyase II (right) catalyzed reactions go through the same carbanion intermediate by abstracting the C_5 proton. This carbanion intermediate can be reprotonated (left) to produce epimerized product, or proceed to an elimination reaction by breaking up the glycosidic bond (right). Left, the C_5 proton is shown in a circle, which can be ^{3}H-labeled.

Several selected tyrosine residues of C_5-epi were mutated to alanine including Tyr-146, Tyr-162, Tyr-168, Tyr-208, Tyr-210, and Tyr-222 followed by analyzing their epimerase activity in comparison to wild type C_5-epi using developed assay. With the exception of Y208A, all of the C_5-epi mutants were successfully expressed in Origami B (DE3) cells as analyzed by SDS-PAGE with comparable level to the wild type enzyme (Fig 41A). The mutants exhibited various levels of C_5-epi activity with both Y168A and Y222A completely losing epimerase activity (Fig 41B). The other tyrosine mutants, Y146A, Y162A, and Y210A, demonstrated a significant level of activity suggesting a less critical role in activity for epimerase.

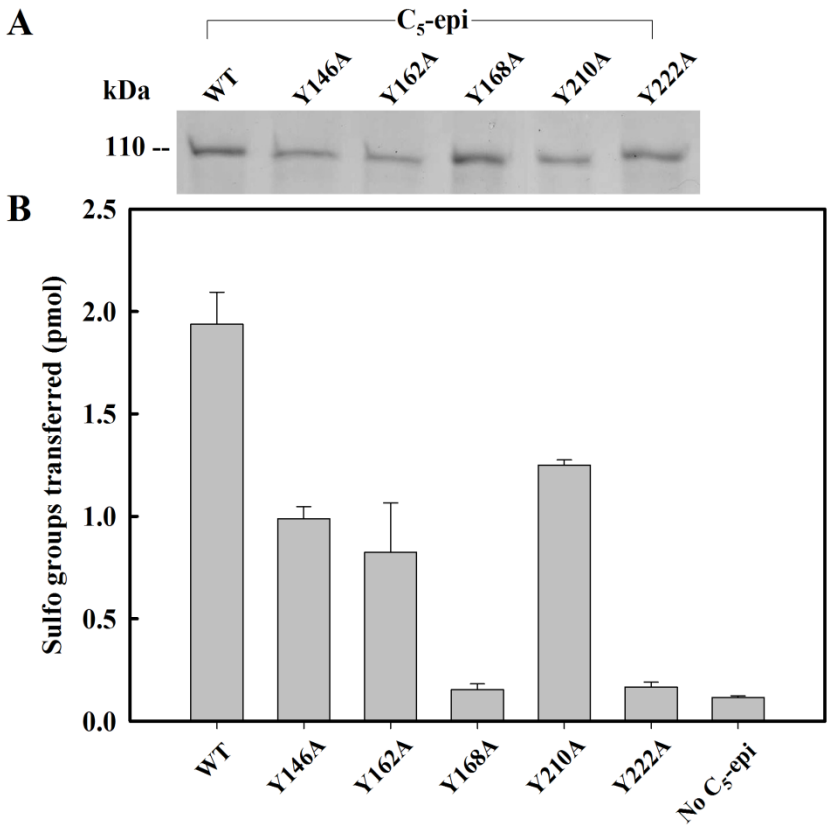


Figure 41. Analysis of tyrosine mutants of C_5-epi. The tyrosine mutants were prepared by site directed mutagenesis using the plasmid containing truncated form of human C_5-epi (Glu^{53}-Asn^{617}) as template. The purified proteins were analyzed by SDS-PAGE, shown in Panel **A**. The activities were shown in Panel **B**.

The results for the C_5-epi Tyr mutants were confirmed by disaccharide analysis of the 2-*O*-sulfated polysaccharide products using an established method (240). The biosynthesis of the IdoUA2S-containing polysaccharide was reconstituted *in vitro* through incubation of *N*-sulfo heparosan with various C_5-epi tyrosine mutants followed by sulfation with wild type 2OST. Wild type 2OST was used to allow for sulfation of both GlcUA and IdoUA units, which cannot be achieved with the 2OST Y94I selective mutant. The resultant polysaccharides were then subjected to nitrous acid degradation at pH 1.5 to yield 2-*O*-sulfated disaccharide products followed by resolving by reverse phase ion pairing-HPLC to

distinguish between IdoUA2S-AnMan and GlcUA2S-AnMan as illustrated in Figure 42E. The presence of the IdoUA2S-AnMan disaccharide product was indicative of the active form of C_5-epi. The use of wild type C_5-epi resulted in the IdoUA2S-AnMan as the major product as expected (Fig 42A). In the absence of C_5-epi, GlcUA2S-AnMan was the only detected disaccharide product (Fig 42D). As the epimerase activity decreased (Y146A), the ratio of IdoUA2S to GlcUA2S diminished (Fig 42B) compared to the wild type. For the Y168A mutant exhibiting no epimerase activity, there was no IdoUA2S-AnMan present (Fig 42C), similar to the disaccharide analysis in the absence of C_5-epi (Fig 42D).

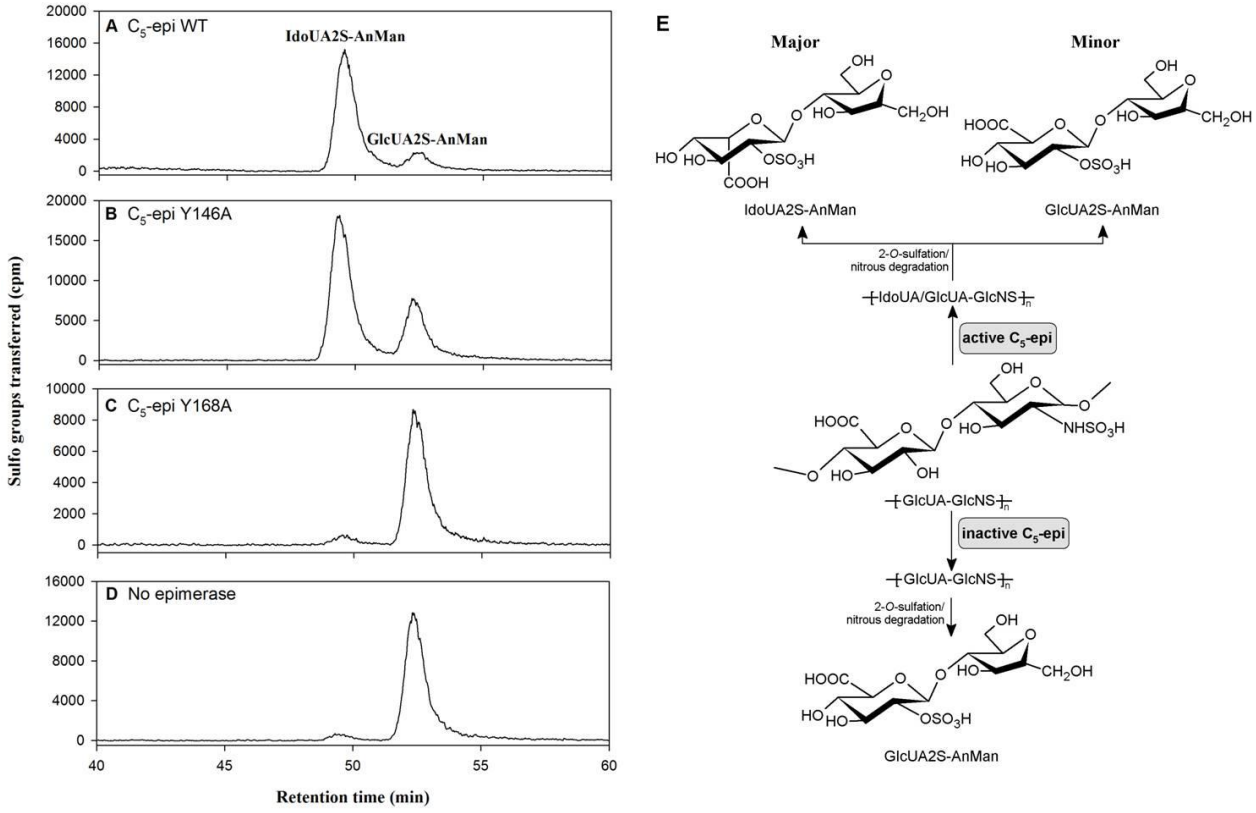


Figure 42. RPIP-HPLC chromatograms of the disaccharide analysis of the polysaccharides modified by various C_5-epi mutants and 2OST WT. The polysaccharides were prepared by incubating with C_5-epi and C_5-epi point mutants followed by the [^{35}S]sulfation using 2OST WT. The preparation of the disaccharides is illustrated in Panel **E**. The resultant disaccharides were resolved using RPIP-HPLC as described under "Experimental Procedures". Panel A-D showed the chromatograms of the polysaccharide treated with **(A)** C_5-epi WT, **(B)** C_5-epi Y146A, **(C)** C_5-epi Y168A and **(D)** no C_5-epi, respectively.

Additional data for the disaccharide analysis of the C_5-epi tyrosine mutants was provided in Table 2. Comparison of the ratio of the two disaccharides and the relative C_5-epi revealed a general trend. As the C_5-epi activity decreased, the ratio of the two disaccharides decreased. These results suggested that the specific tyrosine residues of Tyr-168 and Tyr-222 were essential to the activity of C_5-epi. A crystal structure of C_5-epi would provide a more detailed understanding of the role these two tyrosine residues play in the function of the

enzyme. It is worth noting that the disaccharide analysis was consistent with the results from the newly developed assay, further validating the method.

Table 2. Summary of disaccharide analysis for C_5-epi and 2OST modified HS.

C_5-epi mutants	IdoUA2S-AnMan/GlcUA2S-AnMan	Relative C_5-epi activity[a]
		%
No C_5-epi	≤0.1	--
C_5-epi WT	5.3	100
C_5-epi Y146A	2.2	40
C_5-epi Y162A	2.2	23
C_5-epi Y168A	≤0.1	Inactive
C_5-epi Y210A	6.2	71
C_5-epi Y222A	≤0.1	Inactive

[a]The relative activities were calculated from the data shown in Figure 42.

Conclusions

With the help of our engineered 2OST enzyme identified in Chapter III, a sulfotransferase based assay was developed for measuring the activity of C_5-epi. One of our research goals was to understand the mechanism of C_5-epi however a simple and rapid assay for measuring the activity of this enzyme was currently available. In this study, we developed a sulfotransferase based assay for measuring the C_5-epi activity using the 2OST mutant (Y94A) that specifically sulfates the IdoUA unit. We ensured that the Y94A mutant was capable of detecting the presence of IdoUA in the polysaccharide by analyzing the amount of sulfo groups transferred to either *N*-sulfo heparosan (GlcUA-containing polysaccharide) and CDSNS-heparin (IdoUA-containing polysaccharide) over a range of

2OST concentrations. The Y94A mutant enzyme was incapable of sulfating *N*-sulfo heparosan however it showed excellent activity towards CDSNS-heparin, suggesting that the Y94A mutant could be used to detect and modify solely IdoUA within the polysaccharide.

We first attempted to expand the substrate selectivity of the 2OST Tyr-94 mutant toward IdoUA units by substituting the Tyr-94 with different types of amino acid residues including Y94E, Y94F, Y94G, and Y94I. By comparing the susceptibility of the Tyr-94 mutants to *N*-sulfo heparosan and CDSNS-heparin, we found that 2OST Y94A and 2OST Y94I have comparable substrate selectivity. Compared to the other mutants, Y94A and Y94I have the best substrate selectivity for IdoUA units thus far. The mechanistic role of the isoleucine (or alanine) could be revealed with the co-crystal structure of 2OST and an appropriate oligosaccharide substrate.

The similarity in catalytic transition state between heparin lyases and C_5-epi has been noted since both enzymes involve deprotonation at the C_5 position to yield a carbanion intermediate. This similarity may be due in part to structural homology (245). A recent study on dermatan sulfate epimerase, an enzyme catalyzing conversion of GlcUA to IdoUA of chondroitin sulfate/dermatan sulfate, provided some evidence for this idea (246). Based on the crystal structure, three residues, His-202, Tyr-257, and His-406, are found in the active site of heparin lyase II from *Pedobacter heparinus* and possibly play a role in catalysis (244). Corresponding histidine (His-205 and His-450) and tyrosine (Tyr-261) residues were identified in the dermatan sulfate epimerase by structural modeling and mutation of these residues resulted in complete loss of epimerase activity. Furthermore, the predicted three-dimensional structure of dermatan sulfate epimerase is very similar to the crystal structure of the heparin lyase II (317). Although no obvious amino acid sequence homology between C_5-

epi and heparin lyases was observed, we speculate that the tyrosine residue may also play an important catalytic role for C_5-epi due to the fact that both heparin lyase II and C_5-epi utilize a similar substrate and yield a carbanion intermediate. We selected tyrosine residues to investigate their roles in catalysis. A total of 31 tyrosine residues are present in human C_5-epi. Our attention was focused on the tyrosine residues located between Asn-92 and Val-235 based on the *N*-terminal truncation mutational analysis and amino acid conservation among different species. We postulated that some of the tyrosine residues between Asn-92 and Val-235 might contribute to the enzymatic activity as C_5-epi N92-N617 is active, whereas C_5-epi W203-N617 is inactive. In addition, we found that Glu-105 toVal-235 is the most conserved region in the protein (73% identity) among different species including zebrafish, platypus, mice and humans. Indeed, mutation Tyr-168 and Tyr-222 of C_5-epi completely abolished activity, suggesting their possible roles in the catalytic function of C_5-epi. Due to the inability to assess the overall structure of the mutants, we could not rule out the possibility of the indirect effect of the tyrosine residues on the catalytic function of C_5-epi. Further experimental and structural studies are necessary to determine the precise roles of these residues.

In conclusion, we were able to develop an effective assay for C_5-epi activity by coupling C_5-epi with the engineered 2OST. Previously, a similar method was employed using wild type 2OST, however wild type 2OST sulfates both GlcUA and IdoUA therefore disaccharide analysis was required to identify the presence of the IdoUA2S-AnMan disaccharide product (240). The former method is timely and less quantitative due to the involvement of disaccharide analysis. Our method takes advantage of the engineered 2OST

that only sulfates IdoUA units, eliminating the need for disaccharide analysis. This simple and efficient assay will enhance studying the structural and functional relationship of C_5-epi.

Chapter V

Towards Crystallization of Heparan Sulfate 6-*O*-Sulfotransferase

Heparan sulfate 6-*O*-sulfation is an essential modification for the synthesis of biologically relevant HS structures, particularly for blood coagulation and fibroblast growth factor-fibroblast growth factor receptor mediated signaling (123). The 6OST enzyme plays a key role in developmental processes among different organisms. For *Drosophila* and *C. elegans*, 6-*O*-sulfation is necessary for tracheal development and neural development respectively (109). Studies have also been completed for different vertebrates including zebrafish and mice. Zebrafish deficient in 6OST exhibited poor vasculogenesis as this enzyme appears to control Wnt-mediated signaling pathways (120). Disruption of 6OST within mice resulted in embryonic lethality and abnormal placentation (117). The results of these knockout studies clearly demonstrate an important biological and physiological role for 6-*O*-sulfation within invertebrates and vertebrates.

Previous studies have identified the necessity of 6OST, especially 6OST-1, for the synthesis of anticoagulant heparan sulfate however the mechanism of this enzyme is poorly understood (106). Our ultimate goal is to co-crystallize one or more of the 6OST isoforms with PAP and/or a heparin-like oligosaccharide. This will help gain new insight into the substrate recognition mechanism by the 6OSTs and provide a more elaborate understanding of the HS sulfotransferase mechanism. Our laboratory has successfully prepared several

heparan sulfate enzymes for crystallization including 3-*O*-sulfotransferase isoforms 1, 3, and 5 as well as 2OST, which makes our chances for crystallizing 6OST very feasible (59,63,150,151). Obtaining a robust protein with high solubility and purity is critical for crystallization, which often takes a considerable amount of time and effort when attempting to identify the best expression vector system. As we aimed towards crystallization of 6OST, we utilized a homology model of 6OST based on sequence homology with structurally known sulfotransferases such as 3OST-3. This homology model provided some insight into the role certain amino acid residues may play in catalysis as well as substrate specificity of 6OST. Preliminary mutational analysis of these residues helped provide some insight into the contribution of specific amino acid residues towards substrate specificity, PAPS binding, and catalysis.

Expression and Purification of 6OST Isoforms

To identify the most promising 6OST protein for crystallization, various expression constructs of 6OST-1 and 6OST-3 possessing different fusion tags were prepared as summarized in Table 3. The results of each expression construct will be discussed in the sections to follow.

Table 3. Summary of analyzed 6OST protein expression constructs

6OST Isoform	Species	Fusion Tag	Sequence	Notes
6OST-3	Mouse	His_6	P120-P449	(1) Low expression @ 5 mg/L (2) Poor solubility
6OST-3	Mouse	MBP	P120-P449	(1) 2-fold higher expression compared to His-6OST-3 (2) Heterogeneous population
6OST-1	Chicken	MBP	Y30-W408	(1) Highest activity among 6OST-1 cuts (2) Protein aggregation (3) Heterogeneous population (4) Precipitation with tag removal
6OST-1	Chicken	MBP	H60-W408	(1) Protein aggregation (2) Heterogeneous population (3) Precipitation with tag removal
6OST-1	Chicken	MBP	P68-W408	(1) Protein aggregation (2) Heterogeneous population (3) Precipitation with tag removal
6OST-1	Chicken	MBP	V85-W408	(1) Lowest activity among 6OST-1 cuts (2) Protein aggregation (3) Heterogeneous population (4) Precipitation with tag removal
6OST-1	Chicken	Trx	Y30-W408	(1) Single monomeric population (2) Decreased protein aggregation
6OST-1	Chicken	Trx	Y30-E382	(1) Single monomeric population (2) Decreased protein aggregation (3) 2.5-fold higher expression compared to 6OST-1 (Y30-W408)

Heparan Sulfate 6OST-3

Based on previous success with crystallization of His_6 tagged 3OST-1 derived from mouse (150), we initially attempted crystallization of a His_6 fusion protein containing the catalytic domain (P120- P449) of mouse 6OST-3. However, we discovered it was extremely difficult to upscale the protein expression for His_6-6OST-3 for further biochemical characterization. The maximum amount of protein obtained for His_6-6OST-3 was 5 mg/L total protein, which was not enough for completing isothermal titration calorimetry studies. Therefore, we transitioned to a maltose binding protein (MBP) fusion protein to help improve

the expression and solubility based on previously reported success (232), resulting in 10 mg/L total protein with greater than 80% purity. The fusion protein was prepared by cloning the mouse catalytic domain (P120-P449) into the pMALc2x/TEV vector using XbaI and HindIII restriction enzyme sites followed by transformation into Origami B cells. The protein was isolated from the Origami B cells and purified by an amylose column, then loaded to a gel permeation chromatography (GPC) column for protein property analysis. GPC analysis revealed that most of the protein eluted as aggregated protein (peak 1) in the void volume at 45 mL (peak 1), suggesting that the protein was not properly folded (Fig 43). Based on the standard curve generated in Figure 25 in Chapter III, the calculated molecular weight of MBP-6OST-3 was determined to be 205 kDa. The theoretical molecular weight of MBP-6OST-3 is approximately 70 kDa. Comparison to the calculated molecular weight suggests that the fusion protein exists as a trimer in solution. Peak 3 is consistent with free MBP with a molecular weight of approximately 42 kDa, which was also confirmed by SDS-PAGE analysis.

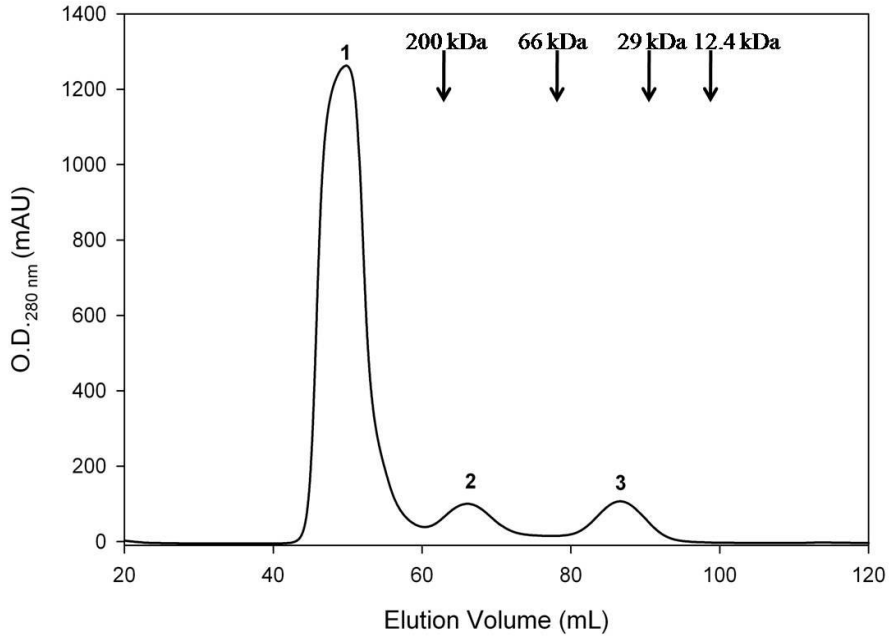


Figure 43. FPLC chromatogram of the purification of MBP-6OST-3 by gel permeation chromatography. Approximately 25 mg of amylose purified MBP-6OST-3 was analyzed by gel permeation chromatography using a Superdex 200 HiLoad 16/60 column. The UV absorbance at 280 nm was plotted against elution volume. Each peak is labeled with a number and identified as follows: 1, aggregated MBP-6OST-3; 2, MBP-6OST-3; 3, MBP.

As seen with 2OST previously discussed in Chapter III, the MBP-2OST protein used for crystallization existed as a trimer in solution which was consistent with the previous results by Kobayashi *et al.*, demonstrating that MBP-2OST isolated from CHO cells migrated as a trimer by GPC (99). Interestingly, the 6OST originally isolated from CHO cells by Habuchi *et al.* migrated as a monomer when analyzed by gel permeation chromatography (108). Based on the CHO cell results, the MBP-6OST-3 protein is not representative of the protein under physiological conditions, suggesting there could be some intermolecular interactions occurring between more than one monomer of MBP-6OST-3. The 6OST protein is known to possess ten cysteine residues, which could potentially allow for disulfide bond

formation among the 6OST monomers. These interactions must be interrupted to heighten the chances of protein crystallization.

Although 6OST-3 is more promiscuous in substrate specificity, the 6OST-1 enzyme is known to be relevant for production of anticoagulant HS (106). Therefore, we decided to attempt crystallization with 6OST-1 to gain understanding of its substrate specificity on the oligosaccharide level as determined by the sulfation patterns surrounding the modification site, specifically for synthesis of the AT-binding pentasaccharide. Previous studies in our laboratory had demonstrated a high level of expression for 6OST-1, offering some promise for crystallization.

Heparan Sulfate 6OST-1

Maltose Binding Protein (MBP) Fusion Protein Expression System

In an attempt to crystallize 6OST-1, the MBP molecule was fused to the *N*-terminus of the catalytic domain of chicken 6OST-1. The 6OST-1 catalytic domain was inserted into pMalc2x vector possessing a TEV protease cleavage site using the restriction enzyme sites of HindIII and BamHI. The TEV protease cleavage site is for removal of the MBP molecule to produce untagged 6OST-1. The higher body temperature for chicken offers a more stable environment for crystallization as was successfully seen with 2OST (59). Several constructs were prepared with different truncations at the *N*-terminus of chicken 6OST-1 as seen in Figure 44A. These *N*-terminal truncations serve to remove the transmembrane region of 6OST-1 with the likelihood of increasing protein solubility.

A.

```
1   MKRAGRTMVERTSKFLLIVAASVCFMLILYQYVGPGLSL
                                    1st cut
40  GAPSGRPYAEEPDLFPTPDPHYVKKYYFPVRELERELAF
                    2nd cut     3rd cut
79  DMKGEDVIVFLHIQKTGGTTFGRHLVQNVRLEVPCDCR
          4th cut
116 PGQKKCTCYRPNRRETWLFSRFSTGWSCGLHADWTELT
155 NCVPGVLGRRESAPNRTPRKFYYITLLRDPVSRYLSEWR
194 HVQRGATWKTSLHMCDGRTPTPEELPSCYEGTDWSGCT
232 LQEFMDCPYNLANNRQVRMLADLSLVGCYNMSFIPENK
270 RAQILLESAKKNLKDMAFFGLTEFQRKTQYLFERTFNLKFI
311 RPFMQYNSTRAGGVEVDNDTIRRIEELNDLDMQLYDYA
349 KDLFQQRYQYKRQLERMEQRIKNREERLLHRSNEALPKE
388 ETEEQGRLPTEDYMSHIIEKW 408
```

B.

N-terminal Cut	Amino Acid Sequence	Oligomeric State (Origami B^{chap})	Oligomeric State (Origami B)
1st cut	Y30-W408	Trimer	Trimer/Monomer
2nd cut	H60-W408	Trimer	Trimer/Monomer
3rd cut	P68-W408	Trimer/Monomer or Timer Only	Trimer/Monomer
4th cut	V85-W408	Trimer/Monomer	Trimer/Monomer

Figure 44. Representation of the *N*-terminal truncations of chicken 6OST-1 and analysis of the oligomeric state using different expression cell lines. (A) Amino acid sequence of chick 6OST-1. The truncation positions of various 6OST-1 constructs are shown on the sequence of chick 6OST-1. **(B)** Protein property analysis of the different MBP-6OST-1 fusion proteins containing various *N*-terminal truncations using Origami B cells with or without the co-expression of chaperonin proteins. Each of the MBP-6OST-1 constructs was purified by amylose column and loaded onto a pre-equilibrated Superdex 200 16/60 HiLoad GPC column for protein property analysis.

The MBP-6OST-1 fusion proteins were expressed in two different Origami B (DE3) cell lines, one of which expresses *E. coli* chaperonin proteins, GroEL and GroES, to assist with protein folding known as Origami B^{chap}. The other Origami B (DE3) cell line does not express the chaperonin proteins. The MBP-6OST-1 fusion proteins were purified by affinity chromatography using an amylose resin followed by purification by a GPC column to determine the oligomeric state.

Figure 45 reflects amylose affinity chromatography (Fig 45A) followed by GPC purification (Fig 45B) of the 3rd *N*-terminal truncation protein containing residues P58-W408

of the 6OST-1 catalytic domain. The GPC chromatograms revealed four distinct peaks (Fig 45B). The size of MBP-6OST-1 is approximately 70 kDa. Based on the standard curve generated in Figure 25 of Chapter III, peaks 2 and 3 were consistent with trimeric and monomeric MBP-6OST-1 respectively. The peaks represent aggregated MBP-6OST-1 (peak 1) in the void volume, trimeric MBP-6OST-1 (peak 2), monomeric MBP-6OST-1 (peak 3), and MBP (peak 4). As with the MBP-6OST-3 protein, most of the protein existed in the aggregated form in the void volume, suggesting improper folding. We clearly demonstrated that the chaperone-assisted MBP-6OST-1 protein expression level was much higher. Several fractions from each peak were analyzed for sulfotransferase activity among the chaperone assisted 3rd cut MBP-6OST-1 protein. Of these peaks, the only ones containing sulfotransferase activity were those consistent with the trimeric (peak 2) and monomeric (peak 3) 6OST-1. The presence of two populations of the 3rd cut 6OST-1 with similar enzymatic activity will decrease the likelihood of obtaining crystals since the protein has the ability to shift between the monomeric and trimeric states in solution. Interestingly, repeating transformation and protein expression of the 6OST-1 3rd cut plasmid into Origami B^{chap} cells resulted in solely a trimeric peak with high sulfotransferase activity when analyzed by gel permeation chromatography. This suggests that the protein has the ability to shift between the monomeric and trimeric states, making it very difficult to isolate the physiologically relevant monomer of 6OST for crystallization.

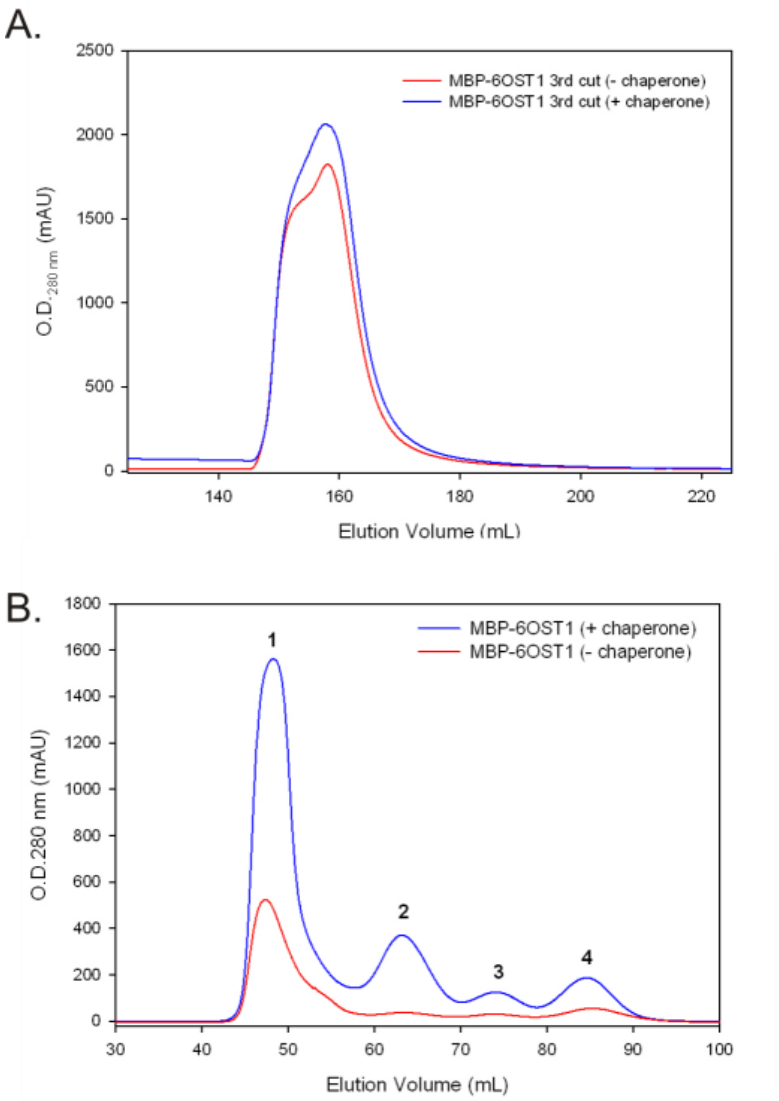


Figure 45. Comparison of differentially expressed chicken MBP-6OST-1 3rd cut fusion protein. The MBP-6OST-1 3rd cut protein (P68-W408) was expressed in two different Origami B (DE3) cell lines, one co-expressing chaperonin proteins to assist with protein folding. The expressed fusion proteins were purified by amylose resin **(A)**, revealing similar expression profiles with (*blue line*) or without (*red line*) the assistance of chaperone. Approximately 25 mg amylose purified MBP-6OST-1with (*blue line*) or without (*red line*) chaperone respectively was analyzed by gel filtration chromatography **(B)**. Each peak is labeled with a number and identified as follows: 1, aggregated MBP-6OST-1; 2, trimeric MBP-6OST-1, monomeric MBP-6OST-1; 4, MBP.

The same experimental procedures were applied to the other *N*-terminally truncated MBP-6OST-1 fusion proteins however GPC analysis revealed conflicting results. Based on the results of the 3rd cut 6OST-1 fusion protein, it would appear that the Origami B^{chap} cells

are the best for improving protein expression, however expression in either cell line resulted in various oligomeric states among all the 6OST-1 different cuts. When expressed in the chaperone assisted Origami B cells, the MBP-6OST-1 proteins existed as trimers with activity solely within the trimer peak. However, expression in the Origami B cells without chaperone assistance resulted in a mixture of both oligomeric states with similar sulfotransferase activity present in both populations.

Based on comparison of the protein property analysis, it appears that the Origami B cell line without chaperone is the best expression cell line for producing the active monomeric 6OST-1 however the major issue is the insubstantial amount of monomer produced during expression. The predominant aggregated 6OST-1 peak (peak 1 in Fig 45B) suggested that that protein is not being properly folded, which could explain the poor expression of the monomer. Protein aggregation was also a major issue for the MBP-6OST-3 protein (Fig 43), which could suggest that the MBP fusion protein is not the best expression system for 6OST. It should be noted that the MBP molecule can be cleaved from 6OST-1 by the use of TEV protease, but cleavage resulted in extensive precipitation. The high level of precipitation implies that the protein possesses poor solubility without the presence of the MBP tag. Therefore, it will be difficult to crystallize 6OST using the MBP expression system due to the major issues of folding, solubility, and oligomerization.

Thioredoxin Fusion Protein Expression System

Studies have indicated that a thioredoxin tag can significantly enhance the solubility of insoluble proteins. Therefore, we turned our attention towards preparation of a thioredoxin-6OST-1 fusion protein (232). The Trx-6OST-1 fusion protein was prepared by performing *N*- and *C*-terminal truncations of chicken 6OST-1 as seen in Figure 47A. To

prepare the *C*-terminally truncated protein, a premature stop codon was introduced at Glu-382 to generate 6OST-1 E382STP, which was previously shown to enhance protein solubility of 3OST-3 (63). The transmembrane (TM) region has been removed from the *N*-terminus of 6OST-1 consistent with the 1st cut presented in Figure 44, since previous studies in our laboratory indicated that this region was not essential for the enzymatic activity. In addition, truncation of the TM domain has been shown to increase protein expression and solubility (247). The *C*-terminal truncation was selected based on a secondary structure prediction using mGenTHREADER that predicted a randomly coiled region at the *C*-terminus of 6OST-1.

The 6OST-1 1st cut DNA was cloned into the pET32b(TEV) vector using the restriction sites BamHI and HindIII, followed by site-directed mutagenesis PCR to generate the premature stop codon at Glu-382. The 6OST-1-pET32b(TEV) plasmid was transformed and expressed in Origami B (DE3) cells, followed by purification by a nickel column coupled to the FPLC. We also prepared a Trx fusion protein containing the 6OST-1 1st cut (Y30-W408) sequence for comparison, which will be referred to as 6OST-1 wild-type (WT). We found that removal of the randomly coiled region at the *C*-terminus of 6OST-1 gave 2-3 fold higher expression while maintaining a full level of sulfotransferase activity.

For a rapid assessment of the oligomeric state for the 6OST-1 WT and E382STP fusion proteins, we utilized gel permeation chromatography coupled to the HPLC. Small scale expression and purification by a benchtop nickel column was completed for both 6OST-1 proteins. A small aliquot of each of the Trx-6OST-1 fusion proteins were then loaded onto a pre-equilibrated G4000SW$_{XL}$ (Tosoh Bioscience) GPC column at a flow rate of 0.5 mL/min using buffer containing 20 mM MOPS pH 7.0 and 400 mM NaCl (Fig 46).

Analysis of the GPC chromatograms revealed that both 6OST-1 WT and E382STP fusion proteins exist as monomers with molecular weights of 58 kDa and 55 kDa respectively. Several fractions from the GPC column were analyzed for sulfotransferase activity and the results revealed that the activity was present solely within the monomeric peak. The sulfotransferase activity was comparable within each sample suggesting that the *C*-terminus is most likely not essential for the 6OST-1 activity.

The thioredoxin tag was cleaved from the fusion proteins to produce untagged 6OST-1 using TEV protease. The cleaved 6OST-1 samples were also analyzed by GPC-HPLC, revealing a similar profile with similar sulfotransferase activity. Unlike the MBP-6OST-1 fusion proteins, the Trx fusion protein can be cleaved to produce a soluble and active monomeric 6OST-1 protein, indicating a potential for crystallization of 6OST-1. Although both 6OST-1 WT and E382STP demonstrated good expression and solubility on a small scale, this may not be indicative of the situation on a large scale. With that said, we must proceed with scaling up the expression of these two enzymes to hopefully determine if one or both may offer promising results for crystallization trials.

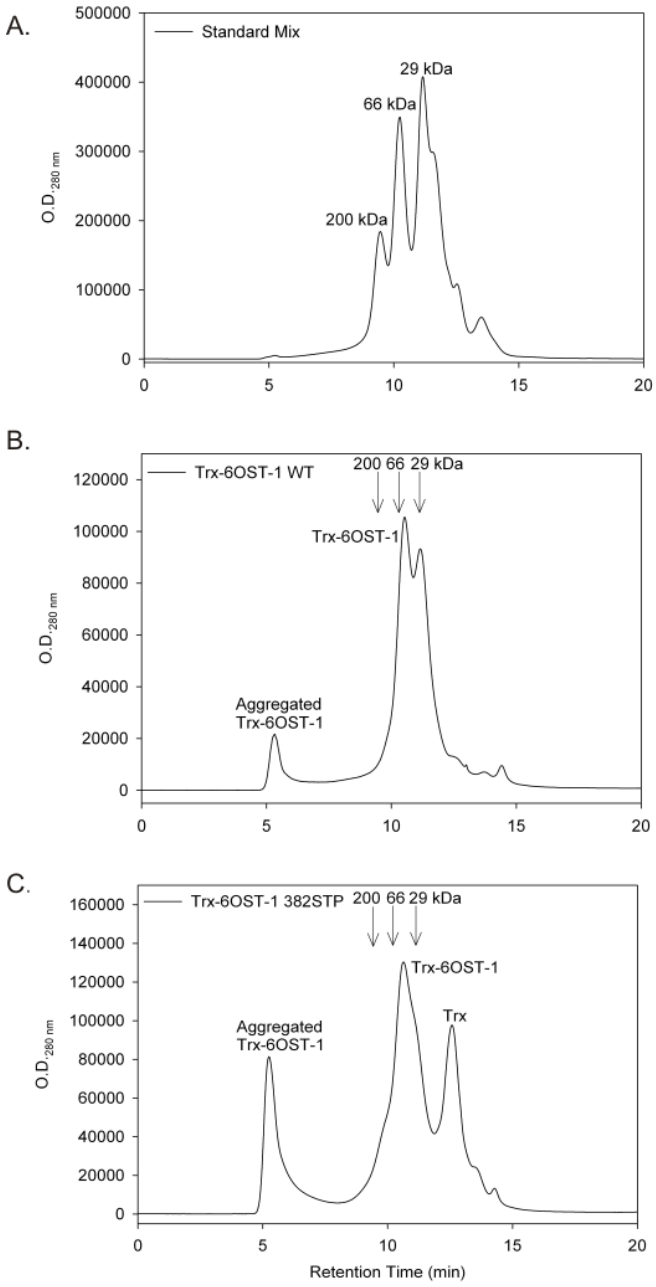


Figure 46. Protein property analysis of Trx-6OST-1 WT and E382STP by GPC-HPLC. All samples were loaded onto a G4000SW$_{XL}$ (Tosoh Bioscience) GPC column equilibrated with 20 mM MOPS pH 7.0 and 400 mM NaCl at a flow rate of 0.5 mL/min. For each Trx-6OST-1 fusion protein, a 250 µL of the nickel purified protein was loaded onto the GPC and the elution was monitored with UV absorbance at 280 nm. **(A)** GPC-HPLC analysis of standard mix. A 50 µL aliquot of the standard mix (Sigma) was loaded and the molecular weight of the protein standards is labeled above each peak. **(B)** GPC-HPLC analysis of Trx-6OST-1 wild type. **(C)** GPC-HPLC analysis of Trx-6OST-1 E382STP. The identity of each protein peak is labeled and the elution times of the standards are indicated by an arrow.

Upon completion of the GPC-HPLC, the Trx-6OST-1 WT and E382STP fusion proteins were expressed in Origami B cells and purified on a large scale (3-liter prep) as outlined in Figure 47C. The cell extract was first loaded onto a nickel-agarose column. Once eluted from the nickel column, the thioredoxin tag was cleaved from the 6OST-1 protein using Tobacco etch virus (TEV) protease followed by two additional purification steps, the first of which included binding heparin-sepharose resin (New England Biolabs) followed by gel permeation chromatography (GPC) using a HR 16/60 Superdex 75 (Amersham) column. Approximately 0.13 mg and 0.32 mg of total protein of WT and E382STP 6OST-1 respectively remained after several rounds of purification, suggesting the expression level is very low. SDS-PAGE analysis of 6OST-1 E382STP following each round of purification revealed a significant increase in purity (Fig 47B). It may appear that only 50% of the protein was recovered after GPC purification seen in Lane 3, however, the heparin-purified sample loaded on the gel had been concentrated for other experimental studies. There was greater than 90% of enzymatic activity recovered after each round of purification. The SDS-PAGE analysis following each round of purification is not shown for the WT 6OST-1 protein however the results were similar with a substantial amount of sulfotransferase activity present solely within the monomeric peak.

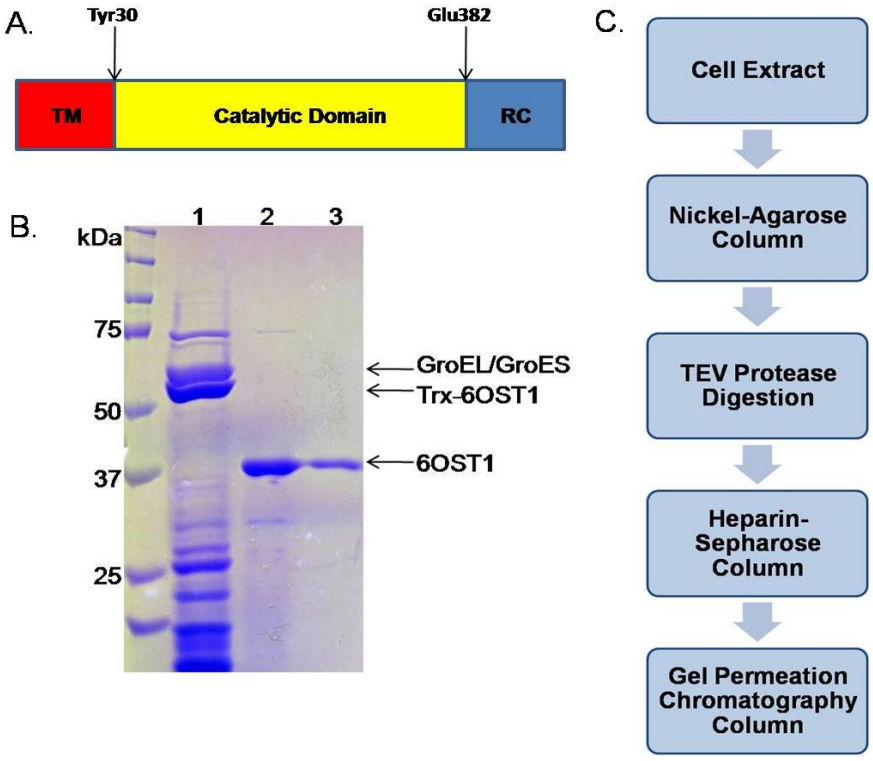


Figure 47. Purification of chicken 6OST-1 E382STP. (A) Domain structure of chicken 6OST-1. The *N*-terminal and *C*-terminal truncations are located at Tyr30 and E382 respectively. TM represents transmembrane region; RC represents random coiled region. **(B)** Flow chart outlining the steps towards purification of 6OST-1 E382STP. **(C)** SDS-PAGE analysis of 6OST-1 E382STP following each round of purification. Lane 1, Nickel-purified Trx-6OST-1 prior to TEV digestion; Lane 2, Heparin-purified 6OST-1 following TEV digestion; Lane 3, GPC-purified 6OST-1 following heparin purification.

The solubility of the cleaved 6OST-1 WT and E382STP proteins was then examined following GPC purification. The two proteins were dialyzed against a buffer containing 25 mM Tris pH 7.5 and 150 mM NaCl to reduce the salt concentration. Ideally, the protein must exist in a low salt environment up to a maximum of 200 mM NaCl to allow hydrogen bond formation between protein molecules. However, lowering the salt concentration can often result in protein precipitation. Therefore, we wanted to drive the limits of the WT and E382STP 6OST-1 proteins to ensure protein stability. The protein concentration was monitored by UV_{280nm} following dialysis for the WT and E382STP, giving concentrations of 0.014 mg/mL (0.13 mg total) and 0.03 mg/mL (0.32 mg) respectively. Based on comparison

of the GPC profiles and amount of total 6OST-1 protein obtained, the expression of E382STP appears to be 2.5 fold higher than that of the wild type protein. Thus, we decided to pursue crystallization studies using 6OST-1 E382STP.

In order to obtain a substantial amount of untagged 6OST-1 E382STP, the level of expression of the Trx fusion protein must be considerably increased. We decided to increase the level of bacterial expression from 3-liter to 10-liter. The untagged 6OST-1 E382STP was prepared as presented in Figure 47C. As was previously demonstrated by the rapid GPC-HPLC analysis, analysis by gel permeation chromatography coupled to FPLC revealed a monomeric 6OST-1 E382STP protein (43 kDa) eluted at 60 mL with little aggregated protein at 45 mL (Fig 48A). We then wanted to test the purity and solubility of the GPC-purified 6OST-1 E382STP protein. The protein was concentrated up to 16-fold at a concentration of approximately 1 mg/mL with constant monitoring by SDS-PAGE after each concentration step as seen in Figure 48B. There appears to be no loss in protein during each concentration step, suggesting that the protein is highly soluble and there is no protein precipitation. Taken together, our data suggest that we should be able to obtain a soluble 6OST-1 protein with exceptionally high purity and enzymatic activity. The limiting factor is the amount of material recovered after the purification steps. After 16x concentration, the 6OST-1 E382STP protein was only at 0.6 mg/mL with a total amount of approximately 600 µg from a 10-liter preparation. Despite the low yield, we continued our pursuit to crystallizing 6OST-1 E382STP due to the presence of a highly pure, soluble, and active monomeric 6OST-1 protein that are all essential protein qualities for crystallization.

The expression was scaled up to a 36-liter prep to enhance amount of protein, obtaining 37.5 mg (1.5 mg/mL) of Trx-6OST-1 E382STP protein from the nickel column.

Following TEV protease cleavage and the heparin and GPC column purification steps, approximately 3.4 mg pure cleaved 6OST-1 E382STP was isolated. There was clearly a significant loss of material during the steps towards purification, which could easily be lost during the TEV cleavage step if the material began to precipitate as only the supernatant is applied to the heparin column.

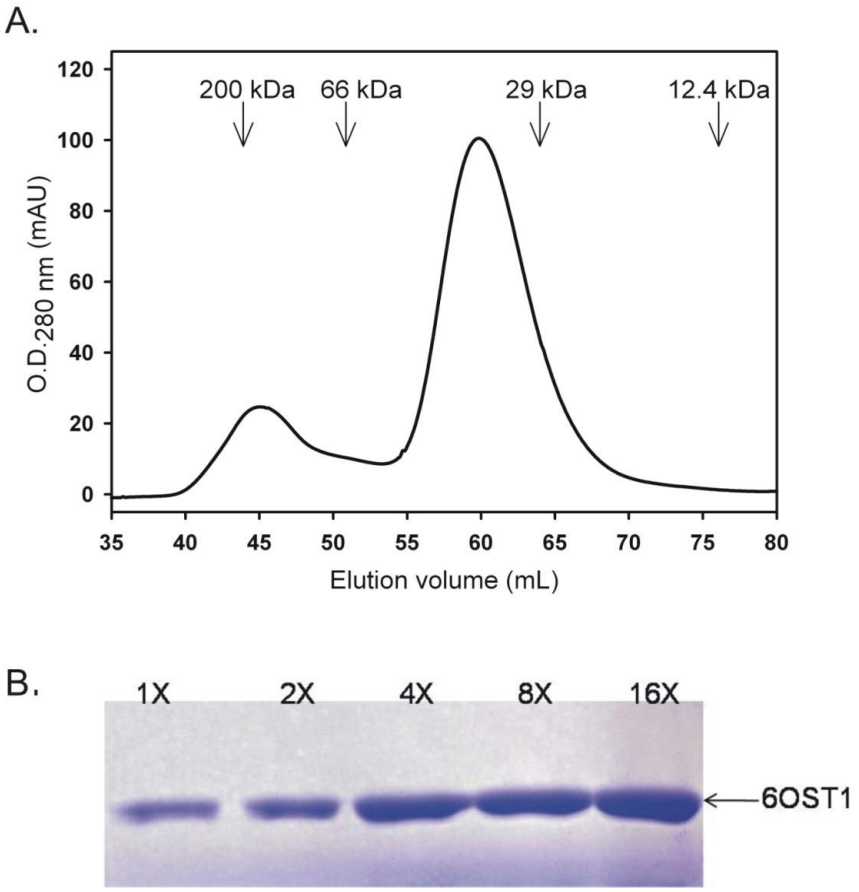


Figure 48. Determination of the oligomeric state of 6OST-1 E382STP. (A) Approximately 8 mg of heparin-Sepharose column purified 6OST-1 E382STP was analyzed by gel permeation chromatography using a Hi-Load Superdex 75 column at a flow rate of 1 mL/min. The UV absorbance at 280 nm is plotted against elution volume. The molecular mass standards resolved under the same experimental conditions are indicated by arrows. **(B)** SDS-PAGE analysis of GPC-purified 6OST-1 E382STP at different concentrations. The sample was concentrated 16-fold by Centricon (MWCO 10,000) to approximately 1 mg/mL.

The GPC purified 6OST-1 E382STP was concentrated 16X to a final concentration of approximately 2 mg/mL and a small 200 µL aliquot was analyzed for solubility within different salt environments. The 200 µL of protein was split into 100 µL each, currently present in 20 mM MOPS pH 7.0 and 400 mM NaCl, and then diluted to a final concentration of either 100 mM or 200 mM NaCl in 20 mM Tris pH 7.0. Following dilution, each pilot protein sample was concentrated using a Microcon Ultracel YM-10 concentrator and monitored by SDS-PAGE (Fig 49). The pilot sample in 200 mM NaCl could not be concentrated up to 16X due to an issue with the microcon concentrators. However, it is clear that the intensity of the protein band does not increase proportionally within the 200 mM NaCl environment. For the 100 mM NaCl pilot sample, the protein was successfully concentrated to 8X the original concentration however there appears to be a loss in material at the 16X concentration step (Fig 49). Interestingly, there was no obvious protein precipitation during the concentration steps for the 100 mM pilot sample, but there is definitely potential for damaging the resin when applying and removing such low protein volumes to the microcon tubes. At this point, one can assume that 6OST-1 E382STP can exist in extremely low salt environments and still be sufficiently stable. The remaining 16X concentrated GPC purified 6OST-1 E382STP was diluted to a final salt concentration of 100 mM and concentrated to approximately 4 mg/mL in 750 µL. The protein was distributed to our collaborative crystallographer at NIEHS, Dr. Lars Pedersen, to perform crystallization trials but unfortunately there was not a substantial amount of protein for extensive crystallization studies to be successfully employed.

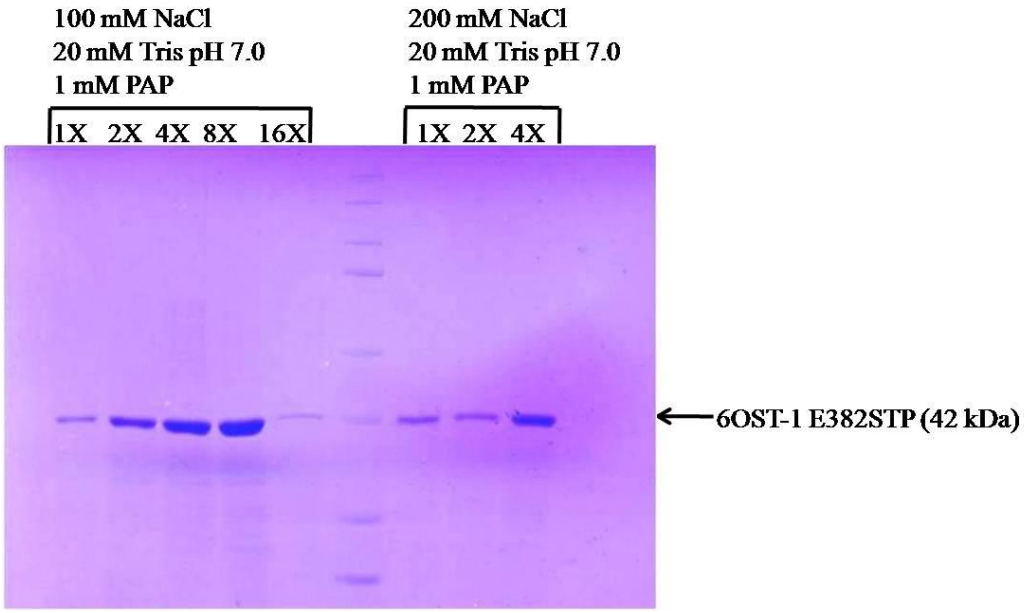


Figure 49. SDS-PAGE analysis of pilot samples of 6OST-1 E382STP within different salt environments. Two pilot samples, 100 µL each, of untagged GPC pure 6OST-1 E382STP were diluted in buffers containing 20 mM Tris pH 7.0 to reduce the salt concentration to 100 mM NaCl or 200 mM NaCl. PAP was added to a final concentration of 1 mM to keep the protein in solution. The pilot samples were concentrated with Microcon Ultracel concentrators as indicated above. The band at 42 kDa represents 6OST-1 E382STP as indicated by the arrow.

Based on the results from the different expression systems utilized for 6OST-1, it appears that the use of the thioredoxin protein expression system offers the most potential for isolation of the physiologically relevant monomeric 6OST-1 protein with considerable sulfotransferase activity. The expression of the premature stop codon E382STP is significantly higher than the WT 6OST-1 however it still poses the major drawback of a low yield for completion of crystallization studies. In the near future, we aim to clone the 6OST-1 E382STP into a fixed arm pET32bX vector, which possesses similar function to the pMALX vector used for crystallization of 2OST. As with the pMALX vector, the linker region of pET32bX vector encodes several alanine residues to reduce the flex in the fusion protein (59). This will allow the 6OST-1 protein to tightly fold next to the thioredoxin protein to allow for crystallization. No TEV cleavage site is present within the pET32bX

vector and therefore the Trx-6OST-1 protein will be crystallized as a fusion protein. The combined use of the thioredoxin protein expression system and Origami B cells for bacterial expression should serve to enhance the solubility and folding of 6OST-1 in the cytoplasm thus enhancing the likelihood of a crystal structure for 6OST.

Mutational Analysis of 6OST-3

A structural neighbor prediction using GenThreader identified both 3OST-3 and NST-1 as the closest in fold recognition to that of 6OST. We generated a homology model for mouse 6OST-3 based on the crystal structure of 3OST-3 (Fig 50). From the amino acid sequence alignment of 6OST-3 and 3OST-3, we identified several amino acid residues that could play a potential role in PAPS binding, substrate recognition and catalytic function. These amino acid residues were systematically mutated to alanine and analyzed for sulfotransferase activity using CDSNS-heparin as seen in Table 4. Several of the residues are evolutionarily conserved across species, suggesting their central role in the function of 6OST.

Table 4. Preliminary mutational analysis of heparan sulfate 6OST-3

HS 6OST-3 Mutants	Sulfotransferase Activity (%)	Potential Defect	Evolutionary Conservation
WT	100		
H151A	56	S[a]	Y
H151Y	<1	S	Y
Q153A	22	S	Y
H151A/Q153A	2	S	Y
K154A	<1	P[b]	Y
K181A	100	S	Y
K182A	100	S	N
K190A	83	S	N
K191A	10	S	N
E192A	7	C[c]	N
D210A	52	C	Y
E213A	88	C	Y
R244A	<1	P	Y
S252A	8	P	Y
E253A	47	S	Y
H256A	14	C	Y
R259A	100	S	Y
T381A	74	S	Y
R382A	90	S	N
S384A	76	S	N

The potential defect correlates with the role a certain amino acid residue may play in 6OST function based on structural homology with 3OST-3. The defects are labeled as follows: [a] Binding to polysaccharide substrate; [b] Binding to PAPS; [c] Catalytic base for ST reaction. The conservation of the amino acid residue across various species is indicated by Y (yes) or N (no).

As were determined with the other structurally known cytosol and *Golgi* resident sulfotransferases, the highly conserved 5'-phosphosulfate (PKT/SGTTW/AL) and 3'-phosphate (IT/YV/I/LLRNPA/KDR/VL/AVSYYY/Q) binding sequence motifs were also identified in 6OST-3 (Fig 50B and 51) (235). Mutation of a potential 5'-phosphate binding residue, Lys-154, of 6OST-3 nearly abolished sulfotransferase activity compared to the wild-type 6OST-3. The corresponding lysine within 3OST-1 (Lys-68), 3OST-3 (Lys-162), EST (Lys-48), and NST-1 (Lys-614) have previously been established as playing a pivotal role in

the sulfo transfer by forming a hydrogen bonding network with the 5'-phosphosulfate group of PAPS (63,150,247,248) (63,80,150). Using the binary complex of 3OST-3 and PAP as a model, the corresponding Lys-154 residue is predicted to be within 2.8 Å of the 5'-phosphate of PAP (Fig 51).

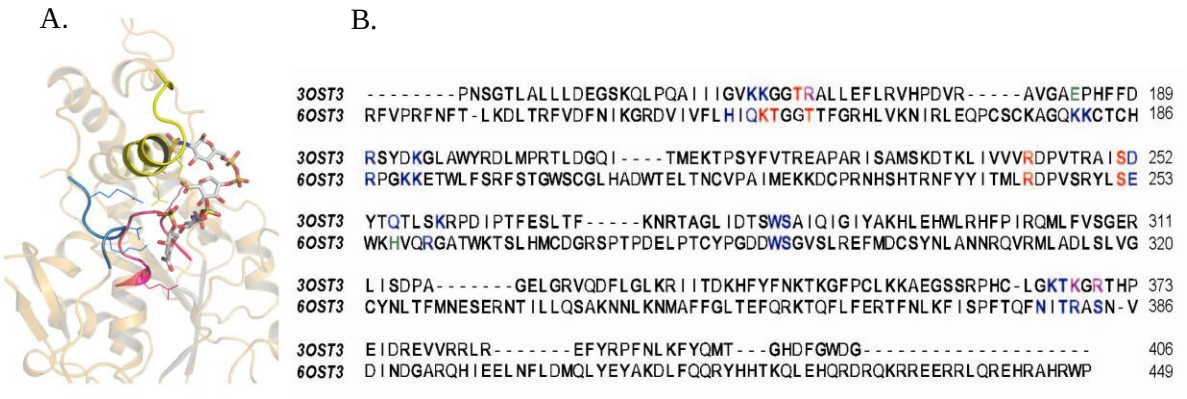


Figure 50. Homology model of 6OST-3. (A) Crystal structure of 3OST-3. The substrate binding sites are highlighted: The PAPS-binding loop is shown in magenta (where Gln^{153} of 6OST-3); the region after helix -5 is shown in yellow (where Arg^{259} of 6OST-3); and the long coil region before a-helix 11 in blue (where Asn^{299}, Thr^{381}, Arg^{382}, and Ser^{384} of 6OST-3). **(B)** Amino acid sequence alignment of 3OST-3 and 6OST-3. The amino acid residues involved in binding to HS are labeled in blue. The amino acid residues involved in binding to PAPS are in red. H256 (in green) is likely to serve as a catalytic base for 6OST. The residues involved in both substrate and PAPS binding are labeled purple.

The potential 3'-phosphate binding residues of Arg-244 and Ser-252 within 6OST-3, which are conserved among all the structurally known sulfotransferases, resulted in 0.6% and 8% sulfotransferase activity respectively when mutated. The serine residue, known to stabilize PAPS in EST, is conserved in NST-1 (Ser-712) and the 3OSTs (Ser-159 in 3OST-1 and Ser-251 in 3OST-3) (249). Based on the binary complex of 3OST-3 with PAP, the corresponding Ser-252 in 6OST-3 can be predicted to be positioned as a part of the conserved α-helix (α5) that is in close proximity to the PAP binding site (Fig 51). With the exception of NST-1, a highly conserved arginine residue exists within the structurally known

HS sulfotransferases and EST that corresponds to Arg-244 in 6OST-3. The preliminary mutational analysis revealed that Lys-154, Arg-244, and Ser-252 of 6OST-3 are most likely involved in binding to the sulfo donor, PAPS, to assist with the sulfuryl transfer reaction. For further verification, these mutants must be analyzed by isothermal titration calorimetry (ITC) for PAP binding as has been established with 3OST-1 and 2OST (104,150). The active donor PAPS is chemically unstable and cannot be utilized during ITC however studies have indicated that PAP binding affinity is indeed representative of PAPS (104). If the mutant proteins do not bind PAP when analyzed by ITC, the loss of sulfotransferase activity can be explained by the decreased PAPS binding. These biochemical characterization studies cannot be completed until the issue of low 6OST expression is resolved.

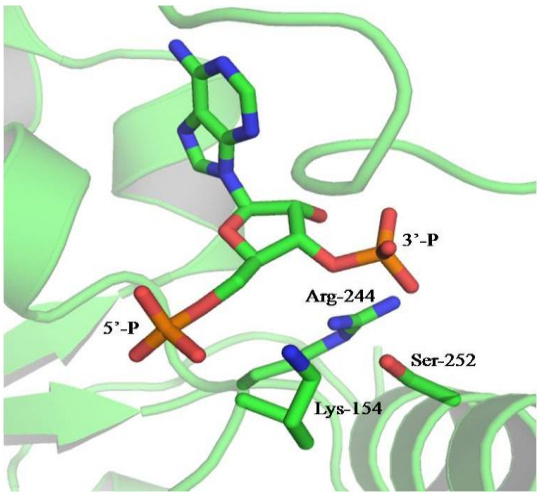


Figure 51. Representation of the PAPS binding residues of 6OST-3 based on the structure of 3OST-3 with PAP bound. With PyMol, the 6OST-3 model was prepared using the structure of the binary complex of 3OST-3 and PAP. The highlighted residues are conserved in both 3OST-3 and 6OST-3 and identified as potentiating PAPS binding. The corresponding 6OST-3 residues are labeled. The Lys-54 of 6OST-3 is predicted to be 2.8 Å away from the 5'-phosphate of PAP. The Ser-252 and Arg-244 residues of 6OST-3 are predicted to be approximately and 2.7 Å and 3.22 Å respectively away from the 3'-phosphate of PAP.

Based on the sequence alignment with 3OST-3, the next subset of 6OST-3 amino acid residues analyzed were those with the potential for HS substrate recognition (Table 4). Several residues were subjected to mutational analysis including His-151, Gln-153, Lys-181, Lys-182, Lys-190, Lys-191, Glu-253, Arg-259, Thr-381, Arg-382, and Ser-284. The only residue showing a noteworthy reduction in sulfotransferase activity was that of Q153A with 22% activity compared to WT. The mutant H151A only saw a reduction in activity to 56% compared to the WT however generation of a double mutation H151A/Q153A saw a strong reduction in activity to 2%. Removal of both residues had a more drastic effect on activity than when only one residue was mutated, suggesting that Gln-153 and His-151 may potentially interact to stabilize the transition state and/or substrate binding. A previous publication identified a single point mutation of H151Y within the 6OST coding domain of the kal-1 gene of *C. elegans*, a gene whose mutation induces a neurological disorder known as Kallman disease (125). Interestingly, generation of the H151Y mutation within 6OST-3 essentially abolished sulfotransferase activity. This result is very difficult to explain as the tyrosine residue would still be in position to form a hydrogen bond with the 5'-sulfate of PAPS as well as interact with the Gln-153, whereas an alanine residue (H151A) loses all these stabilizing interactions. A more extensive mutational analysis must be completed in order to determine the potential influence of these amino acid residues in substrate binding.

Following mutational analysis of potential PAPS and substrate binding residues, the focus turned towards the identification of a catalytic base for 6OST. Based on sequence homology with 3OST-3, several residues were identified as potential catalytic bases, including Glu-192, Asp-210, Glu-213, and His-256. Among these mutant enzymes, only E192A and H256A showed a significant reduction in activity to 7 and 14% activity

respectively (Table 4). The histidine residue is conserved in all 6OSTs among different species, suggesting its importance for the function of 6OST. The glutamic acid residue, on the other hand, is not conserved across species and therefore may not contribute to the catalytic function. The identification of a potential catalytic base for 6OST is inconclusive based on the homology model with 3OST-3 and preliminary mutational analysis. The presence of a crystal structure will serve to identify the contribution of Glu-192 and His-256 to the activity of 6OST.

Conclusions

In order to crystallize 6OST, we had to first identify the best expression system to increase our likelihood of obtaining a highly expressed and soluble protein in a homogeneous population. After several attempts to improve the expression of 6OST-3 using His_6 or MBP tags, we turned our attention to 6OST-1 since it was previously found to possess high expression in our hands. The level of expression and purity of 6OST-1 was greatly enhanced using a MBP fusion tag however the protein exists as a heterogeneous population. Although we can isolate the monomeric form of MBP-6OST-1 by gel permeation chromatography, we are concerned that the monomer can convert to the trimer during crystallization, making crystallization virtually impossible. Ideally, we want 6OST-1 to be present in a single stable population with a high level of sulfotransferase activity. Another issue with the MBP-6OST-1 protein is the presence of aggregated protein in the void volume, indicating a problem with folding. The MBP tag was also incapable of being removed from 6OST-1 without excessive protein precipitation. The drawbacks of a heterogeneous population, improper folding, and protein precipitation upon cleavage suggest that the use of the MBP tag is most likely not the best option for expressing 6OST-1.

Once we exhausted all possibilities with the use of the MBP fusion protein expression system, we focused our attention on the thioredoxin (Trx) fusion protein expression system which is known to improve the solubility and stability of proteins. We worked with two Trx-6OST-1 constructs, WT and the premature stop codon E382STP, identifying both as promoting formation of a highly active monomer when analyzed by GPC-HPLC however the 6OST-1 WT protein had poor expression compared to the premature stop codon. GPC analysis revealed a monomeric 6OST-1 protein with a slight presence of aggregated protein, but it is much lower than that for MBP-6OST-1. At this point, it appears that we have been able to overcome all the issues that were seen with the MBP-6OST-1 fusion protein however the Trx expression system is not without flaws. Although higher compared to 6OST-1 WT, the expression level of the Trx-6OST-1 E382STP protein is still very low. To overcome the issue with low expression, we plan to continue pursuing the Trx expression system but with the use of a fixed arm vector, which will maintain the Trx-6OST-1 as a fusion protein. This fusion protein will have a fixed linker region to reduce flex and allow the 6OST-1 E382STP protein to tuck next to the Trx protein, hopefully promoting crystallization. This fixed arm vector system proved successful with crystallization of the MBP-2OST fusion protein (59).

As we worked towards crystallization of 6OST, we completed preliminary mutational analysis of 6OST-3 based on a sequence homology model with the structurally known 3OST-3 and EST. Several residues were identified that could potentially be involved in PAPS binding and substrate binding. Using the ternary complex of 3OST-3 with PAP and HS tetrasaccharide as a model, Lys-154 was proposed to be located within 3 Å of the 5'-phosphate of PAP while Arg-244 and Ser-252 are potentially within hydrogen bonding distance to the 3'-phosphate of PAP. Based on the combined activity results and homology

model, we can confidently identify these three residues as PAPS binding residues for 6OST-3. The potential substrate binding residues, His-151 and Gln-153, did not reveal a drastic loss in sulfotransferase activity when mutated but a double mutation of both residues resulted in a 50-fold reduction in activity compared to the wild type 6OST-3. These two residues, His-151 and Gln-153, may possibly interact with each other to stabilize the transition state and/or substrate binding to 6OST-1 however this cannot be confirmed without the presence of a ternary complex of 6OST with PAP and an appropriate HS substrate.

Throughout this study, we were capable of identifying a promising expression construct for preparing 6OST-1, leading us one step closer towards crystallization. We also identified several amino acid residues that are important to the function of 6OST. Once the crystal structure of 6OST is solved, we will be able to identify the substrate recognition mechanism and catalytic mechanism of this poorly understood enzyme.

Chapter VI

Conclusions

Heparan sulfate is a highly sulfated linear polysaccharide ubiquitously present on the cell surface and within the extracellular matrix. HS participates in a wide range of physiological and pathophysiological functions, including embryonic development, inflammatory response, blood coagulation, and assisting viral/bacterial infections (250). Heparin, a special form of HS, is a commonly used anticoagulant drug. The sulfation pattern of the HS polysaccharide governs its functional selectivity (251). Understanding the structure–function relationship of HS may allow us to manipulate HS biosynthesis to design HS/heparin with improved anticoagulant efficacy and exploit heparin or heparin-like molecules for the development of anticancer and antiviral drugs (234). However, obtaining HS oligosaccharides or polysaccharides with defined structures remains a challenge. Despite many examples of success with chemical synthesis of short-HS fragments, the synthesis of molecules larger than hexasaccharides is extremely difficult. Using HS biosynthetic enzymes to prepare biologically active polysaccharides and oligosaccharides has offered a promising alternative approach (162,176,240,252).

The HS sulfotransferases contribute to the structural diversity of HS, however the mechanism by which these enzymes recognize the HS substrate and catalyze the transfer of a sulfo group from PAPS to the modification site is poorly understood. Extensive studies for

3OST-3 including the elucidation of a ternary complex of 3OST-3 with PAPS and a HS tetrasaccharide have provided some insight into the general catalytic mechanism of the HS sulfotransferases (63). However, more elaborate work is necessary to understand the substrate recognition mechanism for each sulfotransferase since each is known to possess stringent substrate requirements at and surrounding the modification site. The combination of a structural biology approach with site-directed mutagenesis could prove very effective for elucidating the mechanism of action of the structurally unknown sulfotransferases, 2OST and 6OST.

The initial focus of this dissertation resided within crystallization of 2OST. Over the past few years, several attempts in our laboratory to solve the crystal structure of 2OST had proven unsuccessful due to problems with solubility, heterogeneity, and low expression. However, utilizing a MBP fusion protein eliminated all of these issues at hand and helped us obtain the crystal structure of MBP-2OST in a binary complex with PAP (59). Unlike the other structurally known HS sulfotransferases, 2OST was determined to exist as a trimer that is essential for enzymatic activity. This result is consistent with the 2OST originally isolated from CHO cells, representing the physiologically relevant form of 2OST (99). Interestingly, 2OST possesses an extended *C*-terminal tail that was identified as playing a crucial role in trimer formation as deletion of these *C*-terminal residues disrupted the interactions among the monomers within the trimer.

Using the crystal structure as a guide, several residues along the active site of 2OST were identified that could be involved in substrate binding. The His-142 residue was identified as the proposed catalytic base for 2OST, which contrasts with the NST-1 and the 3OSTs that rely on a glutamic acid as the catalytic base (63,79,81,150,253). This proposed

catalytic residue overlays very well with the catalytic base, His-108, of the cytosolic estrogen sulfotransferase, and alignment of EST with 2OST suggested a higher sequence identity between these two enzymes than with the HS 3OSTs and NST-1. Mutation of His-142 resulted in abolition of sulfotransferase activity, supporting its catalytic role for the function of 2OST. Based on the EST-PAPS structure and preliminary mutational analysis of 2OST, the catalytic residues that coordinate with PAPS were identified and provided better insight into the molecular mechanism of 2OST (80,104). The catalytic mechanism of 2OST is proposed as follows: In the absence of substrate, a hydrogen bonding interaction between 2OST Ser-172 and the 3'-phosphate of PAPS serves to control the dissociation of the 5'-sulfo group of PAPS by Lys-83. The Lys-83 residue coordinates with the side chain oxygen of Ser-172. When the substrate binds in the active site and the His-142 initiates catalysis, the Lys-83 undergoes a conformational change to dissociate from the Ser-172 residue. The Lys-83 then forms a hydrogen bond with the bridging oxygen between the 5'-sulfo and 5'-phosphate groups of PAPS, promoting dissociation of the PAP leaving group. This change in Lys-83 conformation could be caused by the buildup of partial negative charge on the bridging oxygen as the catalytic histidine deprotonates the 2-OH acceptor position of the substrate and then nucleophilically attacks the sulfur atom of PAPS (80). This proposed mechanism could be validated by the presence of a ternary complex of 2OST with an appropriate oligosaccharide substrate.

Further examination of the potential substrate binding residues helped to identify residues that altered the spectrum of specificity for 2OST. Specifically, the mutant R189A only transferred sulfates to GlcUA moieties within the HS polysaccharide whereas 2OST mutants Y94A and H106A preferentially transferred sulfates to IdoUA units. These results

suggest that Arg-189, His-106, and Tyr-94 are important for directing the specificity of 2OST. The exact mechanism is currently unknown due to the lack of an available ternary complex with an appropriate oligosaccharide. It is known that the three selective mutants migrate as trimers when analyzed by gel permeation chromatography, suggesting that these mutations do not support a large scale structural change of 2OST. The difference in specificity is most likely caused by a local effect. Our results demonstrate the feasibility for manipulating the substrate specificity to synthesize selected HS products that cannot be achieved by the wild type 2OST.

Utilizing the substrate selective 2OST mutants, specifically the IdoUA-preferring mutant, we were able to design an assay for the determination of HS C_5-epi activity (239). The current assay for C_5-epi activity includes monitoring the release of a [^{3}H]-proton at the C_5-position of uronic acid of a tritium labeled polysaccharide (93). Although effective, this assay requires an abundant amount of [^{3}H]-labeled water which is not readily prepared in an academic laboratory. The new assay bypasses the need for a radiolabeled polysaccharide through the use of the 2OST mutant. The concept behind the epimerase assay is that a polysaccharide containing solely GlcUA units is incubated with C_5-epi to convert GlcUA to IdoUA. Once the epimerization step is completed, the 2OST Tyr-94 mutant enzyme is added, which will only recognize IdoUA residues within the polysaccharide chain. The level of 2-*O*-sulfation can be monitored by the use of [^{35}S]-labeled PAPS, and this level is proportional to the activity of C_5-epi.

In hopes of increasing the sulfotransferase activity of the Tyr-94 mutant towards the IdoUA-containing substrate, the Tyr-94 residue was subjected to several mutations, including replacing the tyrosine side chain with an acidic or basic residue such as Glu or Arg. The size

of the side chain was altered with the use of Phe or Gly. Based on the results, it appears that only the Y94A and Y94I mutants produced the highest sulfotransferase activity towards the IdoUA-containing substrate. Since Y94A and Y94I possess similar selectivity, the 2OST Y94I mutant was selected for use in the C_5-epi activity assay. The validity of using the 2OST mutant was confirmed by the linear relationship seen with the number of sulfo groups transferred by 2OST Y94I and the C_5-epi concentration. A control experiment using WT 2OST demonstrated that a high level of sulfo groups was incorporated into the GlcUA-containing polysaccharide even in the absence of C_5-epi as WT 2OST is capable of sulfating both GlcUA and IdoUA. This result implies that 2OST WT cannot be used for analyzing the activity of C_5-epi. The C_5-epi activity assay was determined to be useful for screening several C_5-epi mutants differing in truncation size and also helped to identify N92-W203 as being important for the function of C_5-epi.

Based on comparable transition state, the catalytic mechanism of C_5-epi is proposed to be similar to that of the heparin lyases. The presence of a tyrosine residue (Tyr-257) was found to be essential for heparin lyase II catalysis based on the crystal structure of heparin lyase II (244). Although there is no significant sequence homology among heparin lyase II and C_5-epi, we can speculate that the presence of tyrosine residues may also be important for the function of the epimerase as mutation of the tyrosine residue (Tyr-261) in dermatan sulfate epimerase corresponding to Tyr-257 in heparin lyase II abolished activity (246). Based on the C_5-epi truncation data, we located several tyrosine residues for mutational analysis. Confirmed by disaccharide analysis, the tyrosine mutational analysis identified Tyr-168 and Tyr-222 as being essential to C_5-epi function with mutation resulting in complete abolition of activity. These results provide some preliminary insight into the

catalytic mechanism of C_5-epi however a combination of the crystal structure of C_5-epimerase and extensive mutational analysis will help provide more detailed information into the structural mechanism of this enzyme.

Another focus of this research project was geared towards the crystallization of 6OST. As we attempted to identify the ideal expression system for 6OST, we completed preliminary mutational analysis based on a homology model with the structurally known 3OST-3. The fold recognition program, GenThreader, identified 3OST-3 as being the closest in homology to 6OST-3. Based on the sequence alignment between 3OST-3 and 6OST-3, we identified several amino acid residues that may be important for the function of 6OST. Like the other HS sulfotransferases and cytosolic sulfotransferase, EST, 6OST contains the highly conserved PAPS binding motif. Mutation of those residues including Lys-154, Arg-244, and Ser-252 nearly abolished sulfotransferase activity compared to the wild type, suggesting a pivotal role for these residues in recognizing PAPS.

Based on the ternary complex of 3OST-3 with PAP and a HS tetrasaccharide, several residues have been indicated as being essential for binding of the substrate to 3OST-3. These corresponding residues in 6OST-3 identified two residues of Gln-153 and His-151 that may be important for substrate recognition. Single point mutations of these two residues demonstrated no significant reduction in activity, although a double mutant of H151A/Q153A presented a drop in activity to 2%. These results suggest that the residues may interact with one another and PAPS to assist with substrate recognition and/or transition state stabilization. A combination of a crystal structure and extensive mutational analysis should provide a clearer picture of the role of specific amino acids in the function of 6OST.

Alongside mutational analysis, vigorous work was completed to identify the best expression system towards crystallization of 6OST. After several attempts using His_6 and MBP tags which were deemed successful with 3OST-1 and 2OST respectively (59,150), the thioredoxin expression system finally showed the most promise. Two Trx-6OST-1 constructs, WT (Y30-W408) and E382STP (Y30-E382) were prepared, differing in length at the *C*-terminus of 6OST-1. The 6OST-1 E382STP protein is simply a premature stop codon within a randomly coiled region of 6OST-1 based on a secondary structure prediction. The WT and E382STP Trx-6OST-1 fusion proteins were analyzed by GPC-HPLC and determined to exist as monomers, which is consistent with a previous study identifying 6OST isolated from CHO cells as a monomer (108). Although monomeric, the WT 6OST-1 did not exhibit comparable expression to E382STP therefore E382STP was utilized for potential crystallization. Although 6OST-1 E382STP offers promising results, the expression level is still extremely low. The pressing issue with poor expression must be overcome in order to increase the likelihood of obtaining a crystal structure for 6OST-1.

At this point, the thioredoxin expression system has been identified as the most promising system for 6OST-1 expression, and the premature stop codon 6OST-1 E382STP shows the most potential for producing a monomeric protein with potent solubility and activity. To improve the expression level of 6OST-1 E382STP, the E382STP could be cloned into a fixed arm pET32X vector which will produce a Trx-6OST-1 that cannot be cleaved. The fixed arm vector will place the Trx and 6OST-1 proteins in very close contact to reduce protein dynamics and hopefully allow for crystallization. If successful, a more elaborate mutational analysis can be completed to validate the preliminary results and gain further insight into the substrate recognition mechanism of 6OST.

Appendix I

Utilizing Tyrosylprotein Sulfotransferase For the Preparation of Sulfated Proteins

Background & Significance

Tyrosine O-Sulfation

The post-translational modification of the tyrosine residues of several plasma membrane and secretory proteins trafficking the trans-*Golgi* network, mediated by tyrosylprotein sulfotransferase (TPST), has been found to play an essential role in promoting protein-protein interactions between two secretory proteins, a secretory protein and its receptor, and two plasma membrane proteins (254,255). According to the UniProtKB database (http://www.uniprot.org/), there are currently 376 known tyrosine-sulfated proteins, 37 of which are human proteins (256). The majority of the human proteins included blood coagulation proteins, adhesion molecules, peptide hormones, extracellular matrix proteins, and G-protein coupled receptors (Table 5) (256,257).

Table 5. Known human tyrosine sulfated proteins

Class	Protein	Functional Role	Citations
Adhesion Molecules	Endoglycan	L-selectin binding	(258)
	Glycoprotein IBα	Thrombin/vWF binding	(259,260)
	PSGL-1	P- & L-selectin binding	(261-263)
Coagulation Factors	Factor V	Thrombin cleavage	(264,265)
	Factor VIII	Thrombin cleavage, vWF binding	(266,267)
	Factor IX	Unknown	(268-270)
	Fibrinogen γ' chain	Unknown	(271)
Matrix Proteins	Dermatopontin	Unknown	(272)
	Procollagen Type III	Unknown	(273)
	Fibronectin	Unknown	(274)
	Vitronectin	Unknown	(275)
Serpins	α2-antiplasmin	Unknown	(276)
	Heparin cofactor II	Thrombin inhibition	(277)
G-Protein Coupled Receptors	CCR5	CCL3/CCL4/CCL5 binding	(278)
	CCR2B	CCL2 binding	(279)
	CXCR4	CXCL12 binding	(280)
	CX3CR1	CX3CL1 binding	(281)
	C5α receptor	C5a binding	(282)
	TSH receptor	TSH binding	(283)
Gastrin/CCK Family	Gastrin	Prograstrin processing	(284)
	Cholecystokinin	CCK-A receptor binding	(285)
Miscellaneous	α-fetoprotein	Unknown	(286)
	Amyloid precursor protein	Unknown	(287)
	C4 α chain	Cleavage by C1s	(277,288)
	Choriogonadotropin α chain	Unknown	(289)
	FGF-7	Unknown	(290)
	M2B3 antigen	Unknown	(291)
	Thyroglobuliln	Unknown	(292)

Tyrosylprotein sulfotransferases (TPST) catalyze the transfer of a sulfo group from the universal donor PAPS to tyrosine residues of proteins (Fig 52). The detailed catalytic mechanism for the TPST enzymes is currently unknown. Tyrosine sulfation occurs in the trans *Golgi*, preceding the sorting of secretory proteins (293).

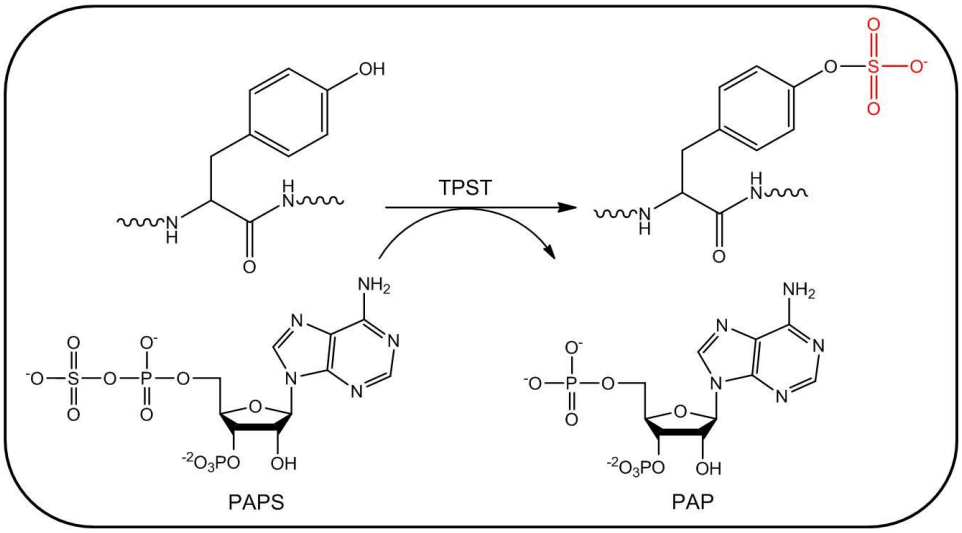


Figure 52. Tyrosylprotein sulfotransferase reaction scheme. TPST catalyzes the transfer of a sulfo group from the universal donor PAPS to the hydroxyl group of a peptidyltyrosine residue to generate sulfotyrosine and PAP.

Two isoforms of TPST exist in both human and mouse, sharing 67% sequence identity between the isoforms. With the exception of *Drosophila malenogaster*, which lacks a second TPST gene, many vertebrates and invertebrates have orthologs for TPST-1 and TPST-2 (256,294,295). The sequence identity among each TPST isoform within different species is greater than 90%, suggesting evolutionary conservation among these enzymes. Both proteins are type II membrane-bound *Golgi*-resident enzymes possessing a short *N*-terminal cytoplasmic tail, a 17-residue transmembrane domain, 40-residue stem region, and catalytic domain containing four conserved cysteine residues (296,297) (Fig 53). Recently, a single TPST gene was identified in the plant *Arabidopsis*, which is strikingly different from TPSTs identified in animal species in that it is a type I transmembrane protein with a larger catalytic domain (446 residues) and more *N*-glycosylation sites possessing no significant sequence homology with sequences of animal species (294,298). The only noteworthy sequence homology is that of residues 371-447 of *Arabidopsis* TPST with a *C*-terminal

region of heparan sulfate 6OST-2. Although the *Arabidopsis* TPST is also located in the *Golgi*, it lacks any 5'-PSB and 3'-PB motifs that are highly conserved within the cytosolic and carbohydrate sulfotransferases (294). It will be interesting to determine if the catalytic mechanism and structures of the plant and animal TPSTs are similar.

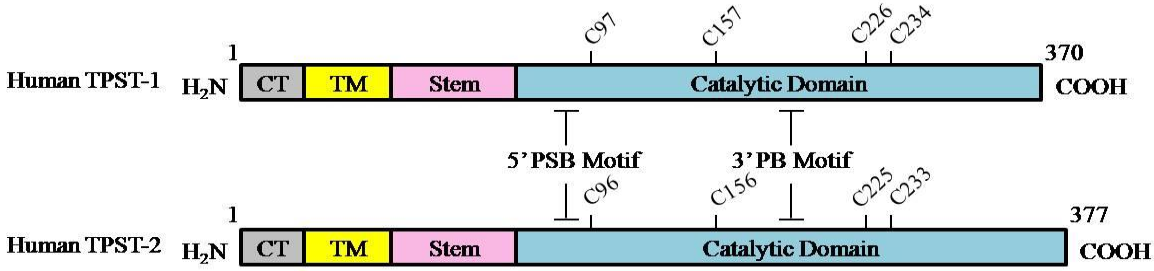


Figure 53. Domain structure for human TPST enzymes. Human TPST-1 and -2 are type II transmembrane proteins of similar size (370-377 amino acid residues) with three domains including a short *N*-terminal cytoplasmic tail (CT), transmembrane (TM) region, stem region, and a *Golgi* luminal catalytic domain. The two isoforms share 67% sequence identity (299). Four cysteine residues including the 5'-PSB and 3'-PB motifs are conserved within the catalytic domain of TPST from all species.

TPST-1 and TPST-2 are optimally active at pH 6.0 and 6.5 respectively and are stimulated by the presence of manganese but inhibited by the presence of calcium. TPST-2 is also stimulated by the presence of magnesium whereas TPST-1 is not (256,299). Previous activity studies with different peptide substrates including P-selectin glycoprotein-1 and several chemokine receptors have identified TPST-1 as having a much lower K_M and V_{max} compared to TPST-2, implying higher enzymatic activity for TPST-1 (299). These enzymatic properties are very important to note for future studies to be presented. Based on sequence comparisons and *in vitro* studies with various synthetic peptides, a consensus sequence has been predicted for tyrosine sulfation that includes an acidic residue (Glu or Asp) adjacent to the Tyr and at least three acidic amino acids within five residues of Tyr. A basic residue adjacent to the Tyr residue was found to abolish sulfation (293,300,301). The

tyrosine residue must be significantly separated from glycosylation sites and cysteine residues, both of which have been found to prevent TPST binding (293,302).

The TPST enzymes are expressed in all human tissues however the level of expression was varied in various tissues. TPST-1 is highly expressed in the reproductive organs including testis and uterus, in neuronal tissues including brain cerebellum, fetal brain, and spinal cord (256,299). On the other hand, TPST-2 is highly expressed in the liver, bladder, blood, bone marrow, thyroid, placenta, and other various organs (256). The distinct expression levels of these two isoforms suggest that these enzymes possess distinct biological functions and different protein target specificities (256).

The role of tyrosine *O*-sulfation in animal development was examined through *in vivo* studies using mice. Mice deficient in TPST-1 exhibited reduced body weight and increased postimplantation fetal death (303) whereas deficiency in TPST-2 lead to infertility in males (297). The reduced body weight in TPST-1 deficient mice has been speculated to be due to the lack of tyrosine sulfation of two known sulfated gastrointestinal peptide hormones known as gastrin and CCK (303). A double knockout study in mice demonstrated death in the early postnatal period due to cardiopulmonary insufficiency. Those mice that were able to survive the postnatal period exhibited primary hypothyroidism shortly thereafter, suggesting that certain tyrosine sulfated proteins were required for pulmonary function at birth and normal thyroid function postnatally (304). A compensatory effect has been suggested for TPST activity since deletion of a single TPST gene did not affect survival rate as was the case for deletion of both TPST genes. These results suggest a significant role for tyrosine *O*-sulfation in distinct physiological functions.

Tyrosine *O*-sulfation plays a role in several physiological and pathophysiological functions, some of which include inflammation, pathogen invasion, and hemostasis. One of the most well-studied tyrosine sulfated proteins is that of P-selectin glycoprotein ligand-1 (PSGL-1), a protein involved in leukocyte-mediated inflammation. One hallmark of inflammation is the migration of leukocytes into tissues. Leukocytes will roll on and adhere to endothelial cells lining blood vessel walls at a site of inflammation. This adhesion step is mediated by the interaction between P-selectin on activated endothelial cells with PSGL-1 present on leukocytes. Extensive studies have indicated the importance of the N-terminal 20 amino acid residues of PSGL-1 in P-selectin binding with deletion of these residues resulting in complete abolition of P-selectin binding. There are three potential tyrosine sulfation sites within the *N*-terminal region of PSGL-1 as seen in Figure 54.

Figure 54. The *N*-terminal 20 amino acid residues of mature PSGL-1. The three potential sulfated tyrosine residues are highlighted in red. The acidic amino acid residues are underlined.

The importance of tyrosine sulfation of the three tyrosine residues within the *N*-terminal region of PSGL-1 has been extensive studied. Synthetic peptides similar to the *N*-terminal region of PSGL-1 were found to be successfully sulfated by both TPST-1 and -2 (256,305). Wilkins *et al.* (261) determined that [^{35}S]-sulfate can be metabolically incorporated into PSGL-1, with base hydrolysis causing release of [^{35}S]-sulfotyrosine and a significant reduction in P-selectin binding. These results suggest the feasibility of tyrosine sulfation of the amino terminal tyrosine residues of mature PSGL-1. Simultaneous mutation

of the three tyrosine residues within the 20 amino acid *N*-terminal region of PSGL-1 significantly reduced P-selectin binding (262). In addition, enzymatic desulfation of PSGL-1 using a bacterial arylsulfatase released sulfate from tyrosine and demonstrated a significant reduction in the ability of PSGL-1 to bind P-selectin (261).

Tyrosine sulfation has been found to play an important role in chemokine receptor binding and function for leukocyte trafficking. Chemokines are released from several tissues in response to inflammation and function through the binding and activation of chemokine receptors, G-protein coupled receptors that are expressed on the surface of leukocytes (256). The *N*-terminal region of chemokine receptors, containing tyrosine residues, has been found to be important for interaction with cognate chemokines. For example, the chemokine receptor CXCR3, which has two potential tyrosine sulfation sites, was found to bind with its cognate chemokines, CXCL10/IP-10 and CXCL11/I-TAC, with high affinity however mutation of each tyrosine residue resulted in complete to 5-10 fold reduction of activity compared to wild type (306). Mutation of both tyrosine residues within the CXCR3 receptor displayed no binding with CXCL0/IP-10 or CXCL11/I-TAC (306). Colvin *et al.* also studied the level of [^{35}S]-sulfate incorporation into wild type and single tyrosine mutant CXCR3 receptors. The single tyrosine receptor mutants demonstrated a reduction in [^{35}S]-sulfated incorporation while the double tyrosine receptor mutant exhibited no binding (307)

Although tyrosine sulfation of chemokine receptors was found to be important for leukocyte trafficking during inflammation, studies have also indicated a pathophysiological role for tyrosine sulfated chemokine receptors in pathogen invasion, specifically for HIV infection. Two known chemokine receptors, CCR5 and CXCR4, were determined to act as co-receptors along with CD4 for HIV viral fusion and entry (308-311). The role of tyrosine

sulfation for CCR5 in binding to HIV has been established. Studies have indicated that the *N*-terminal region of CCR5 possesses four potential tyrosine sulfate sites, which are crucial for the entry of HIV-1 into host cells (312). Inhibition of tyrosine sulfation of CCR5 resulted in a significantly reduced binding affinity of CCR5 for gp120/CD4 complexes where gp120 corresponds to a HIV envelope protein that initially binds to CD4 present on host cells (278). *In vitro* binding studies between gp120/CD4 complexes and a synthetic peptide representing the *N*-terminal region of CCR5 revealed the importance of tyrosine sulfation for binding (313). There is only one tyrosine residue present in the *N*-terminal region of CXCR4 however it does not seem to play a major role in HIV entry compared to CCR5 (280). Tyrosine sulfation has also been implicated in stimulating the interaction of a blood antigen known as DARC (Duffy Antigen and Receptor for Chemokines) with the erthyrocyte receptor for the malarial parasite *Plasmodium vivax* (314). The ability of tyrosine sulfated receptors to stimulate invading pathogens such as HIV and malaria somewhat discredits the importance of post-translational modifications in human diseases, but could prove to be potentially successful therapeutic targets for these diseases.

Several blood coagulation proteins including factors V, VIII, IX, and fibrinogen are tyrosine sulfated post-translationally in mammalian cells. Factors VIII and V function to enhance the rate of activation of factor X and prothrombin respectively. Both coagulation proteins possess six potential tyrosine sulfate sites (264,266,267,315). For both factors, the tyrosine residues must be sulfated for efficient thrombin cleavage and full procoagulant activity (267,316). A deficiency in factor VIII results in hemophilia A, the most common type of blood clotting disorder. Recombinant factor VIII derived from Chinese hamster ovary cells has been utilized as a replacement therapy in hemophilia patients (317). Plasma

and recombinant factor VIII possess 100% tyrosine sulfation as well as similar clotting activity, recovery and half-life, making recombinant factor VIII an effective therapeutic for hemophilia A (317).

Factor IX, which serves to activate factor X, undergoes several post-translational modifications during synthesis in mammalian cells including the addition of *N*-and *O*-linked carbohydrates, glutamic acid γ-carboxylation, tyrosine sulfation, serine phosphorylation, and aspartic acid β-hydroxylation (264,317-320). Factor IX is initially synthesized as a 55 kDa protein which then undergoes these extensive post-translational modifications during transit that are essential for the normal functioning of factor IX (269). During activation by cleavage using factor XIa, factor IX is cleaved at two argininyl peptide bonds results in the formation of activation domain (12 kDa) and activated factor IXa (43 kDa)(321). Regarding tyrosine sulfation, there is solely one site for potential tyrosine sulfation at Tyr-155 located within the activation domain of factor IX (269,322). Recombinant factor IX, commonly used during replacement therapy since it is free of human infectious agents, has only minor differences in post translational modifications from plasma-derived factor IX (317). However, the tyrosine residue is less than 15% sulfated in recombinant factor IX, which is speculated to be the cause for approximately 30% reduced recovery in hemophilia patients (269). Due to the reduced recovery, patients must receive a much higher dose (1.2 times the calculated dose) of recombinant factor IX (269). If the issues of recovery of rFIX can be resolved, rFIX could prove to be a very effective therapeutic for Hemophilia B patients. The use of TPST to increase the level of tyrosine sulfation of rFIX could potentially enhance the recovery upon administration.

Materials & Methods

Expression and Purification of MBP-TPST Fusion Constructs

The maltose binding protein (MBP)-TPST fusion protein constructs were created using the same modified pMAL-c2x vector (New England Biolabs) utilized for the MBP-2OST protein. The TPST-pMALX plasmid was transformed into Origami B cells expressing the GroEL/GroES chaperonin proteins as aforementioned. A 1 L stock of LB supplemented with 100 µg/mL carbenicillin, 50 µg/mL kanamycin, 35 µg/mL chloramphenicol, and 12.5 µg/mL tetracycline was inoculated with 50 µL glycerol cell stock of TPST-pMALX and grown overnight at 37°C with constant shaking at 275 rpm. A 60 mL aliquot of overnight cell culture was added to each of the 12 L cultures supplemented with all four antibiotics at the appropriate concentrations. The cultures were grown at 37°C until the $O.D._{600}$ reached 0.7-0.8 following by reduction of the incubator temperature to 23°C. L-arabinose (1 mg/mL) was first added to induce chaperone expression followed by the addition of 0.5 mM IPTG to induce protein expression. The cultures were allowed to shake overnight at 23°C. The cells were pelleted at 6000 rpm for 15 min followed by resuspension in 15 mL buffer A containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT. The cells were sonicated 4 times for 20 seconds each followed by centrifugation at 18000 rpm for 35 minutes. The supernatant was purified by amylose resin, eluting with buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT, and 40 mM maltose. The eluate concentrated down to 10 mL using a Centricon-Plus 70 concentrator with monitoring of the A_{280} followed by concentration with Amicon Ultra-15 concentrator until the volume was approximately 4.5 mL. The concentrated MBP-TPST protein was loaded onto a Superdex 200 16/60 and eluted with buffer containing 20 mM MOPS, 500 mM NaCl, 1 mM DTT, 40 mM maltose pH 7.0. The purified protein was

diluted with heparin buffer A containing 25 mM Tris pH 7.5, 250 mM NaCl, 1 mM DTT, and 5 mM maltose to reduce to NaCl concentration to approximately 250 mM. The diluted material was then loaded onto a heparin column and eluted with buffer containing 25 mM Tris pH 7.5, 750 mM NaCl, 1 mM DTT, and 5 mM maltose. The protein is diluted and stored in 25 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT, 5 mM maltose, and 4 mM PAP. These constructs were prepared by Dr. Raj Gosavi at the National Institute of Environmental Health Sciences.

Expression and Purification of GST-TPST Fusion Constructs

The GST-TPST fusion protein constructs were created using the pGEXT4T3(TEV) vector which contains a TEV protease cleavage site for removal of the GST tag. The TPST-pGEXT4T3(TEV) plasmid was transformed into Origami B cells expressing the GroEL/GroES chaperonin proteins as aforementioned. A 1 L stock of LB supplemented with 100 µg/mL carbenicillin, 50 µg/mL kanamycin, 35 µg/mL chloramphenicol, and 12.5 µg/mL tetracycline was inoculated with 50 µL glycerol cell stock of TPST- pGEXT4T3(TEV) and grown overnight at 37°C with constant shaking at 275 rpm. A 60 mL aliquot of overnight cell culture was added to each of the 12 L cultures supplemented with all four antibiotics at the appropriate concentrations. The cultures were grown at 37°C until the $O.D._{600}$ reached 0.7-0.8 following by reduction of the incubator temperature to 23°C. L-arabinose (1 mg/mL) was first added to induce chaperone expression followed by the addition of 0.5 mM IPTG to induce protein expression. The cultures were allowed to shake overnight at 23°C. The cells were pelleted at 6000 rpm for 15 min followed by resuspension in 15 mL buffer A containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT. The cells were sonicated 4 times for 20 seconds each followed by centrifugation at 18000 rpm for 35 minutes. The supernatant was

purified by glutathione sepharose 4B resin, eluting with buffer containing 25 mM Tris pH 7.5, 500 mM NaCl, 1 mM DTT, and 40 mM glutathione. The eluate was concentrated down to 3.5 mL using an Amicon Ultra-15. The concentrated GST-TPST protein was split into 2 tubes followed by the addition of 50 µL TEV protease to each tube and rocked overnight at 4°C for complete cleavage of the GST tag. The cleaved material was loaded onto a Superdex 200 16/60 column and eluted with buffer containing 20 mM MOPS, 500 mM NaCl, 1 mM DTT pH 7.0. The GPC purified TPST protein was diluted with heparin buffer containing 25 mM Tris pH 7.5, 250 mM NaCl, 1 mM DTT to reduce the final salt concentration for proper binding to the heparin column. The diluted TPST was then loaded onto a heparin column, eluting with 25 mM Tris pH 7.5, 750 mM NaCl, 1 mM DTT. The pooled heparin fractions were then diluted to bring the final NaCl concentration to 200 mM and the protein was stored in 25 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT, 4 mM PAP. These constructs were prepared by Dr. Raj Gosavi at the National Institute of Environmental Health Sciences.

Determination of the Activity of TPST

Based on an established method (323), the rate of sulfation of the synthetic PSGL-1 peptide (QATEYEYLDYDFLPEC) (Global Peptide Services) by tyrosylprotein sulfotransferase (TPST) was assayed by using the following 100 µL reaction conditions: 2 µM ($2x10^5$ cpm) PAPS, 4.5 µg peptide substrate, 40 mM MES pH 7.0, 20 mM $MnCl_2$, 50 mM NaF, 1 mM 5'-AMP, 0.5% Triton X-100, and 100 µg TPST enzyme. The reaction was incubated at 30°C for 20 min followed by termination with 400 µL ice cold 75 mM EDTA pH 7.0. The [^{35}S]-sulfated peptide product was separated from [^{35}S]PAPS by reverse phase chromatography using C18 Sep-Pak cartridges (Waters). The cartridges were washed with 5 mL methanol containing 0.05% trifluoroacetic acid followed by 5 mL water with 0.05%

trifluroacetic acid. Once the sample was loaded onto the pre-equilibrated cartridges, the resin was washed with 6 mL water/0.05% TFA and eluted with 2 mL methanol/0.05% TFA. The amount of [^{35}S]-sulfated peptide was determined by liquid scintillation counting of the total methanol eluent.

SDS-PAGE Analysis of BeneFIX® Coupled to Autoradiography

For a 50 µL reaction, the following components are used: $2x10^5$ cpm [^{35}S]PAPS, 4.5 µg BeneFIX, 40 mM MES pH 7.0, 20 mM $MnCl_2$, 50 mM NaF, 1 mM 5'-AMP, 0.5% Triton X-100, and 100 µg TPST enzyme. The reaction was incubated at 30°C for 20 min followed by the addition of SDS gel sample buffer (1:1 v/v) (Bio-Rad). A 20 µL aliquot of the sample was loaded onto a 10% Tris-HCl SDS gel and run slowly at 50 volts. For protein size comparison, a 5 µL aliquot of [^{14}C]methylated protein ladder (Amersham) subjected to electrophoresis alongside the samples. At the end of electrophoresis, the SDS gel was then soaked in a fixing solution containing 50% methanol and 10% acetic acid followed by application to QuickDraw blotting paper (Sigma) and covered with Saran wrap. The gel was dried for 2 hours at 80°C in a slab gel vacuum dryer followed by exposure to BioMax maximum resolution (MR) X-ray film (Kodak) for several days at -80 °C before processing the film.

Research Introduction

Tyrosine *O*-sulfation, catalyzed by tyrosylprotein sulfotransferase (TPST), is a common post-translational modification in multicellular eukaryotic organisms, affecting protein characteristics such as enzymatic activity, protein longevity, and interactions with other proteins (295,296). Protein tyrosine sulfation plays a role in several biological functions including inflammation, hemostasis, and autoimmunity (264,296,324).

We have turned our attention to the role of tyrosine sulfation in hemostasis, particularly of blood coagulation protein factor IX which has a single tyrosine sulfation site at Tyr-155 (322). Factor IX deficiency leads to the onset of the rare bleeding disorder known as hemophilia B. The current treatment option for hemophilia B patients is factor IX replacement therapy. However, the recovery following intravenous infusion is reduced due to the decreased level of tyrosine sulfation. BeneFIX®, a FDA approved recombinant factor IX (rFIX) therapeutic, is of particular interest because it does not possess complete sulfation. It possesses only 15%, at the Tyr-155 residue, possibly contributing to its low recovery following administration (317). This is strikingly different from plasma-derived FIX which has almost 100% sulfation at the site. Our goal is to successfully sulfate the single tyrosine residue (Tyr-155) present within recombinant factor IX using TPST constructs differing in isoform and truncation in hopes of improving its stability for clinical use. In collaboration with Dr. Lars Pedersen at NIEHS, we would also like to crystallize the TPST enzyme to gain insight into its substrate recognition mechanism.

Here, we present preliminary data that includes screening for the TPST construct with the highest activity towards an extensively used peptide substrate. Following the peptide sulfation assay, we identified two TPST constructs possessing the potential for effective tyrosine sulfation of BeneFIX® by way of Western blot analysis. This initial study demonstrates the feasibility of sulfating recombinant factor IX using tyrosylprotein sulfotransferase.

Experimental Results

Determining the Sulfotransferase Activity of Various TPST constructs

Based on previous success with 2OST using MBP fusion protein expression systems, we initially expressed several human TPST constructs containing various *N*- and *C*- terminal truncations as MBP fusion proteins in Origami B^{chap} cells. The truncations were prepared to probe for the catalytic domain of TPST. We aim to determine the shortest possible TPST construct possessing high sulfotransferase activity to increase our likelihood of protein crystallization. Once expressed, these MBP-TPST fusion proteins were purified by amylose resin followed by gel permeation chromatography and analyzed for sulfotransferase activity using a well-studied synthesized peptide (QATEYEYLDYDFLPEC) (Tables 6 and 7). This peptide represents the *N*-terminal residues of the mature P-selectin glycoprotein ligand 1 (PSGL-1), an inflammatory protein known to be sulfated by TPST. PSGL-1 contains three tyrosine sulfate sites that have been shown to be sulfated in mammalian cells (261). We originally tested the MBP-TPST-1 (Table 6) and MBP-TPST-2 (Table 7) constructs at a pH of 7.0, but we later identified through a literature search that the average optimal pH for both TPSTs was approximately 6.25 (299). Therefore, all of the prepared constructs were tested under both pH conditions. For both TPST-1 and TPST-2 constructs, the reactivity towards the PSGL-1 peptide substrate did indeed increase by at least two-fold or more under the lower pH conditions. These results were consistent with a previous study utilizing a similar PSGL-1 peptide substrate (ATEYEYLDYDFL) in that the activity increases approximately two-fold when reducing the pH from 7.0 to 6.25 (299).

None of the MBP-TPST constructs demonstrated significant activity towards the peptide substrate except TPST-1 (S39-Q355), possessing reactivity towards the substrate of

28.5 pmol SO_4^{-2}/4 µg relative activity at pH 6.25. Based on the activity analysis, there was no general trend in terms of truncation size versus sulfotransferase activity however it appeared that the activity was higher for the TPST-1 constructs. The higher activity for TPST-1 was consistent with Mishiro *et al.* which demonstrated that full length TPST-1 had a much lower K_M and V_{max} compared to TPST-2 (299).

Table 6. Analysis of MBP-TPST-1 fusion proteins to PSGL-1 peptide substrate

TPST-1 Constructs	Reactivity @ pH 7.0	Reactivity @ pH 6.25
	pmol SO_4^{-2}/4 µg peptide	pmol SO_4^{-2}/4 µg peptide
G25-E370	1.7	3.6
S39-Q355	4.9	28.5
S39-E370	3.4	10.3
Y65-Q355	0.2	2.4
Y65-E370	0.1	0.4
K67-Q355	0.1	0.8
K67-E370	0.1	0.4
M69-Q355	0.1	0.8
M69-E370	0.1	0.2

* The specific activity of PAP[^{35}S] is 22,000 cpm/pmol.

The lack of a highly active TPST construct, even for the longer constructs containing the majority of the catalytic domain, raised concerns with the use of the MBP fusion protein expression system. The modified pMAL vector utilized for expression of the MBP-TPST fusion proteins possessed a mutated linker region encoding three alanine residues (A368-A370) to reduce flex in the fusion protein to facilitate crystallization. This would allow the TPST protein to nestle against the MBP protein, potentially obstructing access to the active site. This modified vector does not possess a TEV protease cleavage site therefore TPST cannot exist as a free protein.

Table 7. Analysis of MBP-TPST-2 fusion proteins to PSGL-1 peptide substrate

TPST-2 Constructs	Reactivity @ pH 7.0	Reactivity @ pH 6.25
	pmol SO_4^{-2}/4 µg peptide	pmol SO_4^{-2}/4 µg peptide
G25-S377	1.4	5.6
Y64-S377	0.3	9.6
Y64-K354	0.2	3.5
K66-K354	0.1	2.5
K66-S377	0.0	0.3
M68-K354	0.2	1.0
M68-S377	0.1	2.5

* The specific activity of PAP[^{35}S] is 22,000 cpm/pmol.

The results with the MBP-TPST proteins caused us to direct our attention to another fusion tag of glutathione S-transferase (GST). The GST fusion protein system has proven very successful for crystallization of other known HS sulfotransferases including NST-1 and 3OST-3 (63,79). We focused on the more significant *N*-terminal truncation constructs of TPST-1 (M69-Q355) and (M69-E370) and TPST-2 (K66-K354) and (K66-S377) to heighten our chances of protein crystallization. The GST-TPST fusion proteins were expressed in Origami B^{chaperone} cells and purified by a GST column. Once eluted from the GST column, the GST-TPST fusion proteins were cleaved by TEV protease to remove the GST tag followed by purification by gel permeation chromatography and finally heparin chromatography. The TPST enzymes, free of tag, were analyzed for activity using the aforementioned PSGL-1 peptide sulfation assay (Table 8), revealing a significant increase in activity for both TPST isoforms compared to the MBP-TPST fusion proteins. These results confirmed the potential inhibitory effects by the presence of the MBP tag. The TPST-1 constructs, differing in sequence length, possess similar activity with the PSGL-1 peptide

regardless of pH conditions, suggesting potential for crystallization as these two proteins exhibited no issues with protein aggregation upon GST cleavage or low expression. Based on the peptide sulfation assay results, the most promising TPST constructs of TPST-1 (M69-Q355) and TPST-1 (M69-E370) were selected to test the sulfation of recombinant factor IX.

Table 8. Analysis of untagged TPST constructs to PSGL-1 peptide substrate

TPST Constructs	TPST Isoform	Reactivity @ pH 7.0	Reactivity @ pH 6.25
		pmol SO_4^{-2}/4 µg peptide	pmol SO_4^{-2}/4 µg peptide
M69-Q355	1	18.1	28.1
M69-E370	1	17.8	30.9
K66-K354	2	9.6	19.2
K66-S377	2	0.9	4.5

* The specific activity of PAP[^{35}S] is 22,000 cpm/pmol.

Utilizing TPST-1 (M69-Q355) & (M69-E370) For Tyrosine Sulfation of rFIX

Upon identification of two highly active TPST-1 constructs, we then wanted to test their ability to sulfate the single tyrosine residue (Tyr-155) of recombinant factor IX. We completed sulfation analysis by incubating a known rFIX therapeutic, BeneFIX®, with each TPST-1 construct in the presence of [^{35}S]-radiolabeled PAPS followed by SDS-PAGE and exposure to X-ray film (Fig 55). The molecular weight of rFIX is approximately 55 kDa, which was consistent with the band present on the autoradiograph for each TPST-1 construct by comparison to a [^{14}C]-radiolabeled protein ladder. Although the two TPST-1 constructs exhibited similar sulfotransferase activity towards the peptide substrate as seen in Table 8, they do not show comparable activity towards factor IX. The band intensity for rFIX modified by the shorter TPST-1 construct (M69-Q355) was slightly higher, suggesting a

higher level of tyrosine sulfation compared to the longer construct. We did attempt sulfation of rFIX with the two TPST-2 constructs of (K66-K354) and (K66-S377) however the autoradiograph revealed no bands corresponding to radiolabeled factor IX, potentially implying that factor IX is not a substrate for TPST-2. Based on this preliminary data, we can conclude that factor IX is indeed a substrate for TPST-1.

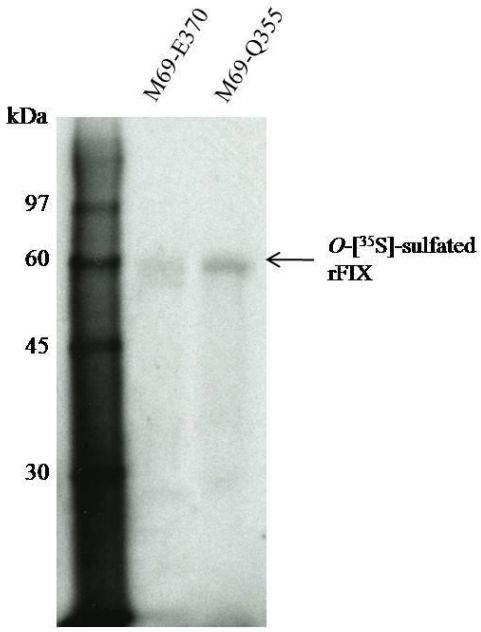


Figure 55. Autoradiograph of recombinant factor IX (BeneFIX®) modified by TPST-1 constructs of various truncation size (M69-E370 and M69-Q355). Approximately 4 µg of BeneFIX® was incubated with 100 µg each TPST-1 enzyme in the presence of $1x10^6$ cpm PAPS followed by SDS-PAGE analysis and exposure to maximum resolution X-ray film. The band at 55 kDa represents *O*- [^{35}S]-sulfated factor IX.

With the recent success modifying rFIX with TPST-1 (M69-Q355), we wanted to test the limit of rFIX saturation by the TPST enzyme. The TPST-1 (M69-Q355) enzyme was incubated with [^{35}S]-radiolabeled PAPS and various concentrations of BeneFIX® ranging from 0-64 µg followed by SDS-PAGE analysis and autoradiography (Fig 56A). As the concentration of rFIX increased, the level of sulfation increased as indicated by an increase in band intensity representing *O*-[^{35}S]-sulfated rFIX at 55 kDa. The dried SDS gel was

exposed to a phosphor screen to quantify the band intensity as the light emitted from the phosphor screen will be proportional to the amount of radioactivity in the sample (Fig 56B). Based on the quantification, it appears that the sulfation of the rFIX substrate reaches the point of saturation at 32 µg. Although there is a decrease in band intensity at 64 µg, we would need to complete further studies to explore the possibility of substrate inhibition. We clearly demonstrated that the level of sulfation by TPST-1 increased as the factor IX substrate concentration increased.

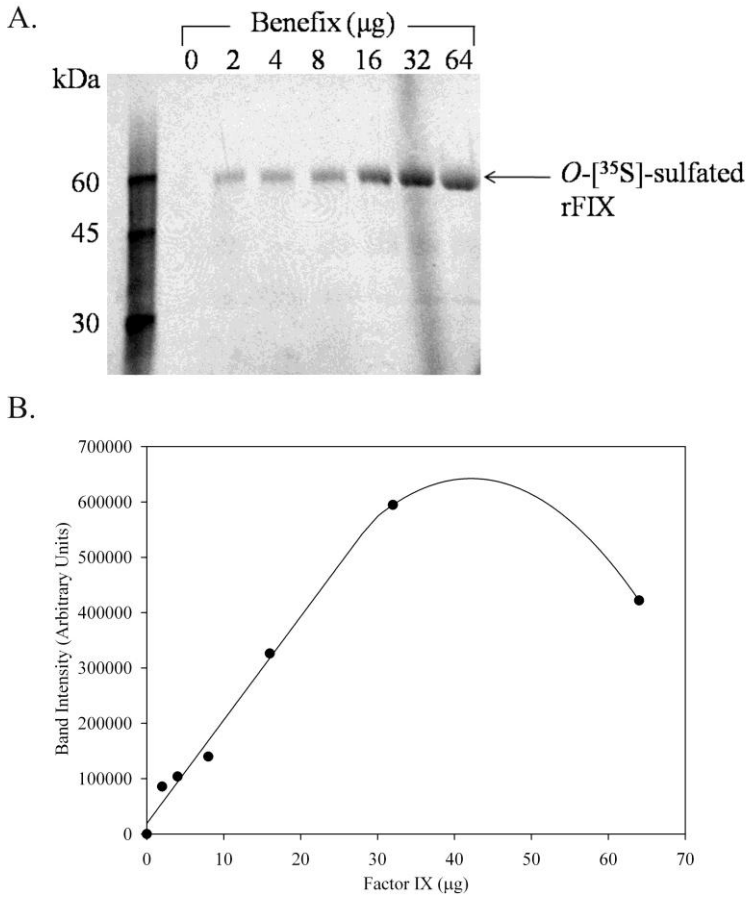


Figure 56. Autoradiograph representing TPST-1 modified recombinant factor IX over various concentrations. (A) Approximately 100 µg of TPST-1 (M69-Q355) was incubated with various concentrations of BeneFIX® (0-64 µg) in the presence of $1x10^6$ cpm PAPS followed by SDS-PAGE analysis and exposure to maximum resolution X-ray film. The band at 55 kDa represents *O*-[^{35}S]-sulfated recombinant factor IX. **(B)** The SDS gel was exposed to a phosphor screen and the bands were quantified using ImageQuant.

Although it was clearly demonstrated that the shorter TPST-1 construct exhibited a higher level of sulfation, we wanted to compare the sulfation efficiency over various rFIX concentrations to determine if the sequence length indeed plays a role in the level of sulfation (Fig 57). Using ImageJ to calculate the band intensity, it does appear that the sulfation efficiency is better for the shorter construct with a proportional increase (2-fold) in band intensity as the concentration of rFIX increased. The longer construct appears to reach saturation more rapidly than the shorter construct, suggesting the binding affinity could be different. Binding studies among the rFIX and TPST-1 constructs using ITC analysis could help explain this difference. Interestingly, the TPST-2 constructs, also differing in sequence length, demonstrated a 4-fold difference in activity when analyzed by the peptide sulfation assay at pH 6.25 (Table 8). Taking this into account along with the discrepancy among factor IX sulfation with the TPST-1 constructs, it appears that the *C*-terminal end of the catalytic domain of TPST could be playing some inhibitory role. We would need to prepare a series of constructs differing in *C*-terminal truncation sites to prove this theory.

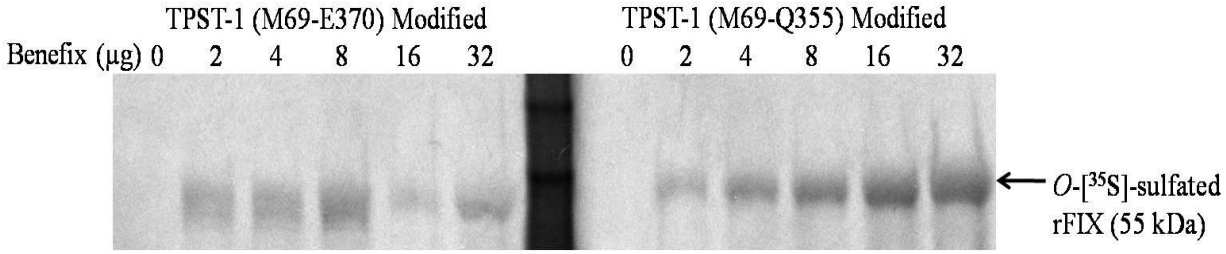


Figure 57. Tyrosine sulfation of recombinant factor IX at various concentrations using two TPST-1 constructs. BeneFIX ranging from 0-32 µg was incubated with [^{35}S]PAPS and approximately 100 µg TPST-1 (M69-E370) or (M69-Q355). Following the ST reaction, the samples were loaded onto a 10% Tris-HCl SDS-gel along with a ^{14}C-labeled protein ladder for size comparison. Following SDS-PAGE, the gel was exposed to an X-ray film and processed after two weeks. The arrow indicates the *O*-[^{35}S]-sulfated factor IX band at 55 kDa.

Conclusions

This research project approaches blood coagulation from a different angle than standard in our laboratory. We normally focus on targeting inhibition of factor Xa and thrombin through binding of heparin/HS to the serine protease inhibitor antithrombin in order to induce anticoagulation. Instead, we turned our attention to promoting coagulation by targeting factor IX within pathway. Hemophilia B, or Christmas disease, is characterized by a deficiency in plasma factor IX. The most common treatment for hemophilia B patients is factor IX replacement therapy using recombinant factor IX derived from bacterial cell systems as with the FDA approved drug, BeneFIX® (317). During the synthesis of Factor IX, the protein is subjected to post-translational processing during transit through the *Golgi*, one of which includes tyrosine *O*-sulfation. Previous studies have indicated that plasma derived factor IX is more than 90% tyrosine sulfated at one position, Tyr-155 (268). Using an effective tyrosine sulfation predictor known as Sulfinator, we confirmed that the Tyr-155 residue was indeed the site of sulfation for factor IX. For recombinant factor IX (rFIX), only 15% of the Tyr-155 residue is sulfated which has been correlated with reduced factor IX recovery, approximately 30% reduction, during treatment (317). Through the use of tyrosylprotein sulfotransferase (TPST) that is known to sulfate tyrosine residues, we could potentially improve the level of sulfation of rFIX and in turn enhance the recovery upon administration.

With intentions to crystallize one of the two TPST isoforms, several constructs were prepared including MBP-TPST and GST-TPST fusion proteins and screened for activity using the peptide substrate, PSGL-1. The construct, such as TPST-1 (M69-Q355), containing the smallest portion of the catalytic domain while still maintaining high

sulfotransferase activity and solubility offers the most promise for crystallization purposes. The screening method was also beneficial for identifying two constructs with potential for sulfating factor IX, including TPST-1 (M69-E370) and (M69-Q355) at the single site of tyrosine sulfation (Tyr-155).

Using SDS-PAGE analysis coupled to autoradiography, we were able to prove that the shorter TPST-1 construct, containing residues M69-Q355, was able to sulfate factor IX over a range of concentrations up to 32 µg whereas the longer TPST-1 construct (M69-E370) appears to demonstrate a significantly lower level of sulfation with saturation at 8 µg factor IX. More studies must be completed to identify any differences in factor IX binding affinity among the two constructs as well as potential inhibition by the substrate. A previously developed assay that relies on mass spectrometry could prove beneficial for understanding the kinetic parameters of these two TPST-1 constructs (325).

Based on the preliminary results, we successfully proved that factor IX is indeed a substrate for TPST-1, although there are some limitations to our current methodology. It appears that the level of sulfation by the TPST-1 enzyme is somewhat low, implying that 100% sulfation may not have been achieved at the site of tyrosine sulfation (Tyr-155). There is the possibility that other cofactors may be involved in promoting tyrosine sulfation of factor IX however there is no current evidence to support this idea. Unlike the peptide substrate, we cannot employ the use of the C18 cartridges for purifying and analyzing tyrosine sulfated FIX as the sulfated protein binds very tightly to the resin and cannot be eluted from the cartridge. Therefore, we must develop an analytical technique for determining the sulfation level of our modified product. We will most likely be able to

utilize RPIP-HPLC analysis. We must select a column resin containing a shorter alkyl chain length such as C4 since we are working with a protein instead of a peptide like with PSGL-1.

Although we are not yet confident that the level of sulfation by the TPST-1 constructs is at 100% for the recombinant factor IX, we have undoubtedly shown that the recombinant TPST-1 enzyme is capable of sulfating factor IX. There is no evidence for utilizing TPST to sulfate factor IX to date, suggesting the novelty of our work. If we are able to prove the sulfation level achieved by the shorter TPST-1 construct, we could potentially complete *in vivo* studies using mice deficient in factor IX to study the implications of the tyrosine sulfated factor IX as an improved replacement therapy.

This project is certainly in the infant stages of development, although our results do offer the possibility of utilizing recombinant enzymes to generate proteins with biological relevance and therapeutic implications.

Appendix II

Curriculum Vitae

Heather Nicole Bethea

Office:
Division of Medicinal Chemistry & Natural Products
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
(919) 962-0065

Home:
1101 Exchange Place Apt 925
Durham, NC 27713
(910) 489-4982

Education

University of North Carolina at Chapel Hill
Ph.D. in Pharmaceutical Sciences
Research Advisor: Associate Professor Jian Liu
Expected May 2010

University of North Carolina at Chapel Hill
B.S. in Chemistry
May 2004

Academic Honors

Dissertation Completion Fellowship
One of 9 awarded by the University of North Carolina at Chapel Hill Graduate School to Ph.D. candidates engaged in an outstanding quality of research, August 2009-May 2010

Graduate School Travel Grant
One of 11 awarded by the University of North Carolina at Chapel Hill Graduate School to selected Ph.D. candidates presenting at national conferences, March 2009.

Off-Campus Dissertation Research Fellowship
Awarded by the University of North Carolina at Chapel Hill Graduate School to Ph.D. candidates conducting research outside the university. January 2009-May 2009

Research Experience

Ph.D. Dissertation Research Assistant, August 2005-Present
University of North Carolina at Chapel Hill, Division of Medicinal Chemistry & Natural Products

Ph.D. Honorary Research Assistant, March 2009-June 2009
University of Liverpool, Department of Biological Sciences, Liverpool, UK
Mentor: Professor Jerry Turnbull

Immunology Research Technician, June 2004-June 2005
Duke University Medical Center, Department of Immunology
Mentor: Professor Youwen He, M.D.

Biochemistry Research Technician, August 2003-May 2004
University of North Carolina at Chapel Hill, Department of Chemistry
Mentor: Distinguished Professor Matthew Redinbo

Teaching Experience

Biochemistry Teaching Assistant, Fall 2005 & Fall 2006
University of North Carolina at Chapel Hill, School of Pharmacy

- Proctored and graded exams, graded problem sets
- Lectured during recitation three times each semester
- Held office hours one hour biweekly

Publications

Li K, **Bethea HN**, Liu J, (2010) Determining the activity of heparan sulfate C_5-epimerase using engineered 2-*O*-Sulfotransferase, Accepted to *J. Biol. Chem*.

Bethea, HN, Xu D, Liu J, Pedersen LC (2008) Redirecting the substrate specificity of heparan sulfate 2-*O*-sulfotransferase by structurally guided mutagenesis. *Proc Natl Acad Sci USA*. 105: 18724-18729.

Draper D, **Bethea H**, He YW (2006) Toll-like receptor 2-dependent and independent activation of macrophages by Group B Streptococci. *Immunol Lett* 102(2): 202-214.

Presentations

<u>Oral</u>

Bethea, H, *Structural and Functional Analysis of Heparan Sulfate 2-O-Sulfotransferase*, National Institute of Environmental Health Sciences, RTP, NC, February 16, 2010.

Bethea H, *Redirecting the Substrate Specificity of Heparan Sulfate 2-O-Sulfotransferase*, 67th Harden Conference - Decoding the biology of heparan sulphate proteoglycans, Robinson College, Cambridge, UK, March 29-April 2, 2009.

Bethea H, *Structural and Functional Analysis of Heparan Sulfate 2-O-Sulfotransferase*, Departmental Seminar, University of North Carolina at Chapel Hill, November 5, 2008.

Bethea H, *A Novel Activator of Procaspase-3 as an Anti-Cancer Strategy*, Departmental Seminar, University of North Carolina at Chapel Hill, April 25, 2007.

Poster

Bethea H, Xu D, Liu J, Pedersen LC, *Redirecting the Substrate Specificity of Heparan Sulfate 2-O-Sulfotransferase by Structurally Guided Mutagenesis*, 67th Harden Conference - Decoding the biology of heparan sulphate proteoglycans, Robinson College, Cambridge, UK, March 29-April 2, 2009.

Leadership Experience

Boka W. Hadzija Award Selection Committee Member, February 2009
University of North Carolina at Chapel Hill Graduate School

Social Chair, August 2005-Present
University of North Carolina at Chapel Hill, Med. Chemistry & Natural Products

Treasurer, August 2007-May 2009
University of North Carolina at Chapel Hill, Graduate Student Organization

REFERENCES

1. Neha, S. G., and Ricardo, L. M. (2008) *Chem. Biol. Drug Des.* **72**(6), 455-482

2. Peterson S, F. A., Liu J. (2009) *Nat. Prod. Rep.* **26**, 610-627

3. Iozzo, R. V. (2001) *J. Clin. Invest.* **108**(2), 165-167

4. Jian, L., and Suzanne, C. T. (2002) *Med. Res. Rev.* **22**(1), 1-25

5. Lever, R., and Page, C. P. (2002) *Nat Rev Drug Discov* **1**(2), 140-148

6. Esko, J. D., and Lindahl, U. (2001) *J. Clin. Inves.* **108**(2), 169-173

7. Toshihiko Toida, H. Y., Hidenao Toyoda, Ichiro Koshiishi, Toshio Imanari, Ronald E. Hileman, Jonathan R. Fromm and Robert J. Linhardt (1997) *Biochem. J.* **322**, 499-506

8. Maccarana, M., Sakura, Y., Tawada, A., Yoshida, K., and Lindahl, U. (1996) *J. Biol. Chem.* **271**(30), 17804-17810

9. Bernfield, M., Gotte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) *Ann. Rev. Biochem.* **68**(1), 729-777

10. Mulloy B, F. M., Jones C, Davies DB. (1993) *Biochem. J.* **293**(849-858)

11. Mulloy, B., and Forster, M. J. (2000) *Glycobiology* **10**(11), 1147-1156

12. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., and Carrell, R. W. (1997) *Proc. Natl. Acad. Sci.* **94**(26), 14683-14688

13. Li, W., Johnson, D. J. D., Esmon, C. T., and Huntington, J. A. (2004) *Nat. Struct. Mol. Biol.* **11**(9), 857-862

14. Raman, R., Venkataraman, G., Ernst, S., Sasisekharan, V., and Sasisekharan, R. (2003) *Proc. Natl. Acad. Sci.USA* **100**(5), 2357-2362

15. Conrad, H. E. (1998) *Heparin-Binding Proteins*, Academic Press, Urbana

16. Ronald, E. H., Jonathan, R. F., John, M. W., and Robert, J. L. (1998) *BioEssays* **20**(2), 156-167

17. Kazuyuki, S., and Hiroshi, K. (2002) *IUBMB Life* **54**(4), 163-175

18. Esko, J. D., and Selleck, S. B. (2002) *Ann. Rev. Biochem.* **71**(1), 435-471

19. Schwartz, N. B. (1977) *J. Bio. Chem.* **252**(18), 6316-6321

20. Kearns, A. E., Campbell, S. C., Westley, J., and Schwartz, N. B. (1991) *Biochem.* **30**(30), 7477-7483

21. Kuhn, J., Götting, C., Schnolzer, M., Kempf, T., Brinkmann, T., and Kleesiek, K. (2001) *J. Biol. Chem.* **276**(7), 4940-4947

22. Schon, S., Prante, C., Bahr, C., Kuhn, J., Kleesiek, K., and Gotting, C. (2006) *J. Biol. Chem.* **281**(20), 14224-14231

23. Götting, C., Kuhn, J., Zahn, R., Brinkmann, T., and Kleesiek, K. (2000) *J. Mol. Biol.* **304**(4), 517-528

24. Esko, J. D., Stewart, T. E., and Taylor, W. H. (1985) *Proc. Natl. Acad. Sci. USA* **82**(10), 3197-3201

25. Casanova, J. C., Kuhn, J., Kleesiek, K., and Götting, C. (2008) *Biochem. Biophys. Res. Commun.* **365**(4), 678-684

26. Brinkmann, T., Weilke, C., and Kleesiek, K. (1997) *J. Biol. Chem.* **272**(17), 11171-11175

27. Bourdon, M. A., Krusius, T., Campbell, S., Schwartz, N. B., and Ruoslahti, E. (1987) *Proc. Natl. Acad. Sci. USA* **84**(10), 3194-3198

28. Esko, J. D., and Zhang, L. (1996) *Curr. Opin. Struct. Biol.* **6**(5), 663-670

29. Sugahara, K., and Kitagawa, H. (2000) *Curr. Opin. Struct. Biol.* **10**(5), 518-527

30. Roch, C., Kuhn, J., Kleesiek, K., and Götting, C. (2010) *Biochem. Biophys. Res. Commun.* **391**(1), 685-691

31. Almeida, R., Levery, S. B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E. P., and Clausen, H. (1999) *J. Biol. Chem.* **274**(37), 26165-26171

32. Bai, X., Zhou, D., Brown, J. R., Crawford, B. E., Hennet, T., and Esko, J. D. (2001) *J. Biol. Chem.* **276**(51), 48189-48195

33. Okajima, T., Yoshida, K., Kondo, T., and Furukawa, K. (1999) *J. Biol. Chem.* **274**(33), 22915-22918

34. Okajima, T., Fukumoto, S., Furukawa, K., Urano, T., and Furukawa, K. (1999) *J. Biol. Chem.* **274**(41), 28841-28844

35. Furukawa, K., and Okajima, T. (2002) *Biochm. Biophys. Acta* **1573**(3), 377-381

36. Kitagawa, H., Tone, Y., Tamura, J.-i., Neumann, K. W., Ogawa, T., Oka, S., Kawasaki, T., and Sugahara, K. (1998) *J. Biol. Chem.* **273**(12), 6615-6618

37. Helting, T., and Roden, L. (1969) *J. Biol. Chem.* **244**(10), 2799-2805

38. Brandt, A. E., Distler, J., and Jourdian, G. W. (1969) *J. Biol. Chem.* **64**(1), 374-380

39. Tone, Y., Pedersen, L. C., Yamamoto, T., Izumikawa, T., Kitagawa, H., Nishihara, J., Tamura, J.-i., Negishi, M., and Sugahara, K. (2008) *J. Biol. Chem.* **283**(24), 16801-16807

40. Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A., and Negishi, M. (2000) *J. Biol. Chem.* **275**(44), 34580-34585

41. Pedersen, L. C., Darden, T. A., and Negishi, M. (2002) *J. Biol. Chem.* **277**(24), 21869-21873

42. Wei, G., Bai, X., Sarkar, A. K., and Esko, J. D. (1999) *J. Biol. Chem.* **274**(12), 7857-7864

43. Fritz, T. A., Gabb, M. M., Wei, G., and Esko, J. D. (1994) *J. Biol. Chem.* **269**(46), 28809-28814

44. Kitagawa, H., Shimakawa, H., and Sugahara, K. (1999) *J. Biol. Chem.* **274**(20), 13933-13937

45. Kim, B.-T., Kitagawa, H., Tamura, J.-i., Saito, T., Kusche-Gullberg, M., Lindahl, U., and Sugahara, K. (2001) *Proc. Natl. Acad. Sci. USA* **98**(13), 7176-7181

46. Norton, W. H. J., Ledin, J., Grandel, H., and Neumann, C. J. (2005) *Development* **132**(22), 4963-4973

47. Lee, J.-S., von der Hardt, S., Rusch, M. A., Stringer, S. E., Stickney, H. L., Talbot, W. S., Geisler, R., Nüsslein-Volhard, C., Selleck, S. B., Chien, C.-B., and Roehl, H. (2004) *Neuron* **44**(6), 947-960

48. Kitagawa, H., Egusa, N., Tamura, J.-i., Kusche-Gullberg, M., Lindahl, U., and Sugahara, K. (2001) *J. Biol. Chem.* **276**(7), 4834-4838

49. Lind, T., Tufaro, F., McCormick, C., Lindahl, U., and Lidholt, K. (1998) *J. Biol. Chem.* **273**(41), 26265-26268

50. Kim, B.-T., Kitagawa, H., Tanaka, J., Tamura, J.-i., and Sugahara, K. (2003) *J. Biol. Chem.* **278**(43), 41618-41623

51. Wuyts, W., Van Hul, W., De Boulle, K., Hendrickx, J., Bakker, E., Vanhoenacker, F., Mollica, F., Lüdecke, H.-J., Sayli, B. S., Pazzaglia, U. E., Mortier, G., Hamel, B., Conrad, E. U., Matsushita, M., Raskind, W. H., and Willems, P. J. (1998) *Am. J. Hum. Genet.* **62**(2), 346-354

52. McCormick, C., Duncan, G., Goutsos, K. T., and Tufaro, F. (2000) *Proc. Natl. Acad. Sci. USA* **97**(2), 668-673

53. Lind, T., Lindahl, U., and Lidholt, K. (1993) *J. Biol. Chem.* **268**(28), 20705-20708

54. Senay C, L. T., Muguruma K, Tone Y, Kitagawa H, Sugahara K, Lidholt K, Lindahl U, Kusche-Gullberg M. (2000) *EMBO Reports* **1**(3), 282-286

55. Lin, X., Wei, G., Shi, Z., Dryer, L., Esko, J. D., Wells, D. E. and Matzuk, M. M. (2000) *Dev. Biol.* **24**, 299-311

56. Stickens, D., Zak, B. M., Rougier, N., Esko, J. D., and Werb, Z. (2005) *Development* **132**(22), 5055-5068

57. Sasisekharan, R., and Venkataraman, G. (2000) *Curr. Opin. Chem. Biol.* **4**(6), 626-631

58. Carlsson, P., Presto, J., Spillmann, D., Lindahl, U., and Kjellen, L. (2008) *J. Biol. Chem.* **283**(29), 20008-20014

59. Bethea, H. N., Xu, D., Liu, J., and Pedersen, L. C. (2008) *Proc. Natl. Acad. Sci.* **105**(48), 18724-18729

60. Avci, F. Y., DeAngelis, P.L., Liu, J., Lindhardt, R.J. (2007) Enzymatic Synthesis of Glycosaminoglycans: Improving on Nature. In: Demchenko, A. V. (ed). *Frontiers in Modern Carbohydrate Chemistry*, Oxford University Press

61. Bhattacharya, R., Townley, R. A., Berry, K. L., and Bulow, H. E. (2009) *J. Cell Sci.* **122**(24), 4492-4504

62. Kamiyama, S., Sasaki, N., Goda, E., Ui-Tei, K., Saigo, K., Narimatsu, H., Jigami, Y., Kannagi, R., Irimura, T., and Nishihara, S. (2006) *J. Biol. Chem.* **281**(16), 10945-10953

63. Moon, A. F., Edavettal, S. C., Krahn, J. M., Munoz, E. M., Negishi, M., Linhardt, R. J., Liu, J., and Pedersen, L. C. (2004) *J. Biol. Chem.* **279**(43), 45185-45193

64. Turnbull, J., Powell, A., and Guimond, S. (2001) *Trends Cell Biol.* **11**(2), 75-82

65. Rong, J., Habuchi, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2001) *Biochemistry* **40**(18), 5548-5555

66. Habuchi, H., Tanaka, M., Habuchi, O., Yoshida, K., Suzuki, H., Ban, K., and Kimata, K. (2000) *J. Biol. Chem.* **275**(4), 2859-2868

67. Kjellen, L. (2003) *Biochem. Soc. Trans.* **31**(2), 340-342

68. Li, J.-P., Gong, F., El Darwish, K., Jalkanen, M., and Lindahl, U. (2001) *J. Biol. Chem.* **276**(23), 20069-20077

69. Shworak, N. W., Liu, J., Petros, L. M., Zhang, L., Kobayashi, M., Copeland, N. G., Jenkins, N. A., and Rosenberg, R. D. (1999) *J. Biol. Chem.* **274**(8), 5170-5184

70. Wei Z, S. S., Ishihara M, Orellana A, and C B Hirschberg. (1993) *Proc. Natl. Acad. Sci. USA* **90**(9), 3885–3888

71. Aikawa, J.-i., Grobe, K., Tsujimoto, M., and Esko, J. D. (2001) *J. Biol. Chem.* **276**(8), 5876-5882

72. Aikawa, J.-i., and Esko, J. D. (1999) *J. Biol. Chem.* **274**(5), 2690-2695

73. Hashimoto, Y., Orellana, A., Gil, G., and Hirschberg, C. B. (1992) *J. Biol. Chem.* **267**(22), 15744-15750

74. Eriksson, I., Sandback, D., Ek, B., Lindahl, U., and Kjellen, L. (1994) *J. Biol. Chem.* **269**(14), 10438-10443

75. Orellana, A., Hirschberg, C. B., Wei, Z., Swiedler, S. J., and Ishihara, M. (1994) *J. Biol. Chem* .**269**(3), 2270-2276

76. Kusche-Gullberg, M., Eriksson, I., Pikas, D. S. c., and Kjellen, L. (1998) *J. Biol. Chem.* **273**(19), 11902-11907

77. Berninsone, P., and Hirschberg, C. B. (1998) *J. Biol. Chem.* **273**(40), 25556-25559

78. Duncan, M. B., Liu, M., Fox, C., and Liu, J. (2006) *Biochem. Biophys. Res. Commun.* **339**(4), 1232-1237

79. Kakuta, Y., Sueyoshi, T., Negishi, M., and Pedersen, L. C. (1999) *J. Biol. Chem.* **274**(16), 10673-10676

80. Pedersen, L. C., Petrotchenko, E., Shevtsov, S., and Negishi, M. (2002) *J. Biol. Chem.* **277**(20), 17928-17932

81. Kakuta Y, L. L., Pedersen LC, Pedersen LG, Negishi M. (2003) *Biochem. Soc. Trans.* **31**(2), 331-334

82. Fan, G., Xiao, L., Cheng, L., Wang, X., Sun, B., and Hu, G. (2000) *FEBS Letters* **467**(1), 7-11

83. Grobe, K., Inatani, M., Pallerla, S. R., Castagnola, J., Yamaguchi, Y., and Esko, J. D. (2005) *Development* **132**(16), 3777-3786

84. Zuberi, R. I., Ge, X. N., Jiang, S., Bahaie, N. S., Kang, B. N., Hosseinkhani, R. M., Frenzel, E. M., Fuster, M. M., Esko, J. D., Rao, S. P., and Sriramarao, P. (2009) *J. Immunol.* **183**(6), 3971-3979

85. Forsberg, E., Pejler, G., Ringvall, M., Lunderius, C., Tomasini-Johansson, B., Kusche-Gullberg, M., Eriksson, I., Ledin, J., Hellman, L., and Kjellen, L. (1999) *Nature* **400**(6746), 773-776

86. Duelli, A., Ronnberg, E., Waern, I., Ringvall, M., Kolset, S. O., and Pejler, G. (2009) *J. Immunol.* **183**(11), 7073-7083

87. Pallerla, S. R., Lawrence, R., Lewejohann, L., Pan, Y., Fischer, T., Schlomann, U., Zhang, X., Esko, J. D., and Grobe, K. (2008) *J. Biol. Chem.* **283**(24), 16885-16894

88. Prihar, H. S., Campbell, P., Feingold, D. S., Jacobsson, I., Jensen, J. W., Lindahl, U., and Roden, L. (1980) *Biochemistry* **19**(3), 495-500

89. Li, JP., Hagner-McWhirter, A., Kjellen, L., Palgi, J., Jalkanen, M., and Lindahl, U. (1997) *J. Biol. Chem.* **272**(44), 28158-28163

90. Murphy, K. J., McLay, N., and Pye, D. A. (2008) *J. Am. Chem. Soc.* **130**(37), 12435-12444

91. Crawford, B. E., Olson, S. K., Esko, J. D., and Pinhal, M. A. S. (2001) *J. Biol. Chem.* **276**(24), 21538-21543

92. Malmstrom, A., Rodén, L., Feingold, D. S., Jacobsson, I., Backstrom, G., and Lindahl, U. (1980) *J. Biol. Chem.* **255**(9), 3878-3883

93. Campbell, P., Hannesson, H. H., Sandback, D., Rodén, L., Lindahl, U., and Li, J. P. (1994) *J. Biol. Chem.* **269**(43), 26953-26958

94. Jacobsson, I., Lindahl, U., Jensen, J. W., Rodén, L., Prihar, H., and Feingold, D. S. (1984) *J. Biol. Chem.* **259**(2), 1056-1063

95. Li, JP., Gong, F., Hagner-McWhirter, A., Forsberg, E., Abrink, M., Kisilevsky, R., Zhang, X., and Lindahl, U. (2003) *J. Biol. Chem.* **278**(31), 28363-28366

96. Pinhal, M. A. S., Smith, B., Olson, S., Aikawa, J., Kimata, K., and Esko, J. D. (2001) *Proc. Natl. Acad. Sci.* **98**(23), 12984-12989

97. Rong, J., Habuchi, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2000) *Biochem. J.* **346**(2), 463-468

98. Rong, J., Habuchi, H., Kimata, K., Lindahl, U. & Kusche-Gullberg, M. (2001) *Biochem. J.* **40**, 5548-5555

99. Kobayashi, M., Habuchi, H., Habuchi, O., Saito, M., and Kimata, K. (1996) *J. Biol. Chem.* **271**(13), 7645-7653

100. Merry, C. L. R., and Wilson, V. A. (2002) *Biochim. Biophys. Acta* **1573**(3), 319-327

101. Kinnunen, T., Huang, Z., Townsend, J., Gatdula, M. M., Brown, J. R., Esko, J. D., and Turnbull, J. E. (2005) *Proc. Natl. Acad. Sci.* **102**(5), 1507-1512

102. Kobayashi, T., Habuchi, H., Tamura, K., Ide, H., and Kimata, K. (2007) *J. Biol. Chem.* **282**(27), 19589-19597

103. Bullock, S. L., Fletcher, J. M., Beddington, R. S. P., and Wilson, V. A. (1998) *Genes Dev.* **12**(12), 1894-1906

104. Xu, D., Song, D., Pedersen, L. C., and Liu, J. (2007) *J. Biol. Chem.* **282**(11), 8356-8367

105. Kreuger, J., Salmivirta, M., Sturiale, L., Gimenez-Gallego, G., and Lindahl, U. (2001) *J. Biol. Chem.* **276**(33), 30744-30752

106. Zhang, L., Beeler, D. L., Lawrence, R., Lech, M., Liu, J., Davis, J. C., Shriver, Z., Sasisekharan, R., and Rosenberg, R. D. (2001) *J. Biol. Chem.* **276**(45), 42311-42321

107. Habuchi, H., Kobayashi, M., and Kimata, K. (1998) *J. Biol. Chem.* **273**(15), 9208-9213

108. Habuchi, H., Habuchi, O., and Kimata, K. (1995) *J. Biol. Chem.* **270**(8), 4172-4179

109. Kamimura, K., Fujise, M., Villa, F., Izumi, S., Habuchi, H., Kimata, K., and Nakato, H. (2001) *J. Biol. Chem.* **276**(20), 17014-17021

110. Turnbull, J., Drummond, K., Huang, Z., Kinnunen, T., Ford-Perriss, M., Murphy, M., and Guimond, S. (2003) *Biochem. Soc. Trans.* **31**(2), 343-348

111. Nogami, K., Suzuki, H., Habuchi, H., Ishiguro, N., Iwata, H., and Kimata, K. (2004) *J. Biol. Chem.* **279**(9), 8219-8229

112. Habuchi, H., Miyake, G., Nogami, K., Kuroiwa, A., Matsuda, Y., Kusche-Gullberg, M., Habuchi, O., Tanaka, M., and Kimata, K. (2003) *Biochem. J.* **371**(1), 131-142

113. Jeff, S., Konstantin, I., and Wellington, V. C. (2004) *Dev. Dyn.* **231**(4), 782-794

114. Habuchi, H., Tanaka, M., Habuchi, O., Yoshida, K., Suzuki, H., Ban, K., and Kimata, K. (2000) *J. Biol. Chem.* **275**(4), 2859-2868

115. Smeds, E., Habuchi, H., Do, A.-T., Hjertson, E., Grundberg, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2003) *Biochem. J.* **372**(2), 371-380

116. Jemth, P., Smeds, E., Do, A.-T., Habuchi, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2003) *J. Biol. Chem.* **278**(27), 24371-24376

117. Habuchi, H., Nagai, N., Sugaya, N., Atsumi, F., Stevens, R. L., and Kimata, K. (2007) *J. Biol. Chem.* **282**(21), 15578-15588

118. Monkley, S. J., Delaney, S. J., Pennisi, D. J., Christiansen, J. H., and Wainwright, B. J. (1996) *Development* **122**(11), 3343-3353

119. Zhang, X., Gaspard, J. P., and Chung, D. C. (2001) *Cancer Res.* **61**(16), 6050-6054

120. Bink, R. J., Habuchi, H., Lele, Z., Dolk, E., Joore, J., Rauch, GJ., Geisler, R., Wilson, S. W., den Hertog, J., Kimata, K., and Zivkovic, D. (2003) *J. Biol. Chem.* **278**(33), 31118-31127

121. Manning, G. a. K., M.A. (1993) *The Development of Drosophila melanogaster* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

122. Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U., and Rapraeger, A. C. (1993) *J. Biol. Chem.* **268**(32), 23906-23914

123. Pye, D. A., Vives, R. R., Turnbull, J. E., Hyde, P., and Gallagher, J. T. (1998) *J. Biol. Chem.* **273**(36), 22936-22942

124. Bülow, H. E., and Hobert, O. (2004) *Neuron* **41**(5), 723-736

125. Bulow, H. E., Berry, K. L., Topper, L. H., Peles, E., and Hobert, O. (2002) *Proc. Natl. Acad. Sci. USA* **99**(9), 6346-6351

126. Xu, D. (2006) Structural and Mutational Analysis of Heparan Sulfate Biosynthetic Enzymes. In. *Medicinal Chemistry & Natural Products*, University of North Carolina at Chapel Hill, Chapel Hill

127. Liu, J., Shworak, N. W., Fritze, L. M. S., Edelberg, J. M., and Rosenberg, R. D. (1996) *J. Biol. Chem.* **271**(43), 27072-27082

128. Shworak, N. W., Shirakawa, M., Colliec-Jouault, S., Liu, J., Mulligan, R. C., Birinyi, L. K., and Rosenberg, R. D. (1994) *J. Biol. Chem.* **269**(40), 24941-24952

129. Colliec-Jouault, S., Shworak, N. W., Liu, J., de Agostini, A. I., and Rosenberg, R. D. (1994) *J. Biol. Chem.* **269**(40), 24953-24958

130. Shukla, D., Liu, J., Blaiklock, P., Shworak, N. W., Bai, X., Esko, J. D., Cohen, G. H., Eisenberg, R. J., Rosenberg, R. D., and Spear, P. G. (1999) *Cell* **99**(1), 13-22

131. Borjigin, J., Deng, J., Sun, X., De Jesus, M., Liu, T., and Wang, M. M. (2003) *J. Biol. Chem.* **278**(18), 16315-16319

132. Miyamoto, K., Asada, K., Fukutomi, T., Okochi, E., Yagi, Y., Hasegawa, T., Asahara, T., Sugimura, T., and Ushijima, T. *Oncogene* **22**(2), 274-280

133. Shworak, N. W., Liu, J., Petros, L. M., Zhang, L., Kobayashi, M., Copeland, N. G., Jenkins, N. A., and Rosenberg, R. D. (1999) *J. Biol. Chem.* **274**(8), 5170-5184

134. Tiwari, V., O'Donnell, C. D., Oh, M.-J., Valyi-Nagy, T., and Shukla, D. (2005) *Biochem. Biophys. Res. Commun.* **338**(2), 930-937

135. Liu, J., Shriver, Z., Blaiklock, P., Yoshida, K., Sasisekharan, R., and Rosenberg, R. D. (1999) *J. Biol. Chem.* **274**(53), 38155-38162

136. Xu D, T. V., Xia G, Clement C, Shukla D, Liu J. (2005) *Biochem. J.* **385**, 451-459

137. Adam, B. C., and Yost, H. J. (2006) *Dev. Dyn.* **235**(12), 3423-3431

138. Kamimura, K., Rhodes, J. M., Ueda, R., McNeely, M., Shukla, D., Kimata, K., Spear, P. G., Shworak, N. W., and Nakato, H. (2004) *J. Cell Biol.* **166**(7), 1069-1079

139. Shworak, N. W., HajMohammadi, S., de Agostini, A. I., and Rosenberg, R. D. (2002) *Glycoconjugate J.* **19**(4), 355-361

140. HajMohammadi, S., Enjyoji, K., Princivalle, M., Christi, P., Lech, M., Beeler, D., Rayburn, H., Schwartz, J. J., Barzegar, S., de Agostini, A. I., Post, M. J., Rosenberg, R. D., and Shworak, N. W. (2003) *J. Clin. Invest.* **111**(7), 989-999

141. Yanada, M., Kojima, T., Ishiguro, K., Nakayama, Y., Yamamoto, K., Matsushita, T., Kadomatsu, K., Nishimura, M., Muramatsu, T., and Saito, H. (2002) *Blood* **99**(7), 2455-2458

142. Ishiguro, K., Kojima, T., Kadomatsu, K., Nakayama, Y., Takagi, A., Suzuki, M., Takeda, N., Ito, M., Yamamoto, K., Matsushita, T., Kusugami, K., Muramatsu, T., and Saito, H. (2000) *J. Clin. Invest.* **106**(7), 873-878

143. Liu, J., Shworak, N. W., Sinay, P., Schwartz, J. J., Zhang, L., Fritze, L. M. S., and Rosenberg, R. D. (1999) *J. Biol. Chem.* **274**(8), 5185-5192

144. O'Donnell, C. D., Tiwari, V., Oh, M.-J., and Shukla, D. (2006) *Virology* **346**(2), 452-459

145. Xia, G., Chen, J., Tiwari, V., Ju, W., Li, J.-P., Malmstrom, A., Shukla, D., and Liu, J. (2002) *J. Biol. Chem.* **277**(40), 37912-37919

146. Wu, Z. L., Lech, M., Beeler, D. L., and Rosenberg, R. D. (2004) *J. Biol. Chem.* **279**(3), 1861-1866

147. Chen, J., and Liu, J. (2005) *Biochim. Biophys. Acta* **1725**(2), 190-200

148. Chen, J., Duncan, M. B., Carrick, K., Pope, R. M., and Liu, J. (2003) *Glycobiology* **13**(11), 785-794

149. Zhang, L., Lawrence, R., Schwartz, J. J., Bai, X., Wei, G., Esko, J. D., and Rosenberg, R. D. (2001) *J. Biol. Chem.* **276**(31), 28806-28813

150. Edavettal, S. C., Lee, K. A., Negishi, M., Linhardt, R. J., Liu, J., and Pedersen, L. C. (2004) *J. Biol. Chem.* **279**(24), 25789-25797

151. Xu, D., Moon, A. F., Song, D., Pedersen, L. C., and Liu, J. (2008) *Nat Chem Biol* **4**(3), 200-202

152. Pedersen, J. L. a. L. C. (2007) *Appl. Microbiol. Microtechnol.* **74**(2), 263-272

153. Prabhakar, V., Capila, I., and Sasisekharan, R. (2009) The Structural Elucidation of Glycosaminoglycans. In. *Glycomics*

154. Gailani, D., and Renne, T. (2007) *Arterioscler Thromb Vasc Biol* **27**(12), 2507-2513

155. Mann, K. G., Butenas, S., and Brummel, K. (2003) *Arterioscler Thromb Vasc Biol* **23**(1), 17-25

156. Whisstock, J. C., Pike, R. N., Jin, L., Skinner, R., Pei, X. Y., Carrell, R. W., and Lesk, A. M. (2000) *J. Mol. Biol.* **301**(5), 1287-1305

157. Becker, R. C. (2004) *J. Thromb. Thrombolysis* **18**(1), 55-58

158. Maurice, P., and Constant, A. A. v. B. (2004) *Angew. Chem. Int. Ed.* **43**(24), 3118-3133

159. de Graffenried, C. L., Laughlin, S. T., Kohler, J. J., and Bertozzi, C. R. (2004) *Proc. Natl. Acad. Sci.* **101**(48), 16715-16720

160. Atha DH, L. J., Petitou M, Rosenberg RD, Choay J. (1985) *Biochemistry* **24**(23), 6723-6729

161. Kuberan, B., Lech, M. Z., Beeler, D. L., Wu, Z. L., and Rosenberg, R. D. (2003) *Nat Biotech* **21**(11), 1343-1346

162. Chen, J., Jones, C. L., and Liu, J. (2007) *Chem. Biol.* **14**(9), 986-993

163. Maurice Petitou, C. A. A. v. B. (2004) *Angew. Chem. Int. Ed.* **43**(24), 3118-3133

164. Oosta GM, G. W., Beeler DL, Rosenberg R. (1981) *Proc Natl Acad Sci USA* **78**, 829-833

165. Olson ST, B. I. (1994) *Semin Thromb Hemost* **20**(4), 373-409

166. Eswarakumar, V. P., Lax, I., and Schlessinger, J. (2005) *Cyt. Growth Factor Rev.* **16**(2), 139-149

167. Itoh, D. M. O. a. N. (2001) *Genome Biol.* **2**(Reviews), 3005

168. Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J., and Mohammadi, M. (2000) *Mol. Cell* **6**(3), 743-750

169. DiGabriele, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J., and Hendrickson, W. A. (1998) *Nature* **393**(6687), 812-817

170. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B., and Blundell, T. L. (2000) *Nature* **407**(6807), 1029-1034

171. Pye, D. A., Vives, R. R., Hyde, P., and Gallagher, J. T. (2000) *Glycobiology* **10**(11), 1183-1192

172. Harmer, N. J., Ilag, L. L., Mulloy, B., Pellegrini, L., Robinson, C. V., and Blundell, T. L. (2004) *J. Mol. Biol.* **339**(4), 821-834

173. Jastrebova, N., Vanwildemeersch, M., Rapraeger, A. C., Gimenez-Gallego, G., Lindahl, U., and Spillmann, D. (2006) *J. Biol. Chem.* **281**(37), 26884-26892

174. Maccarana, M., Casu, B., and Lindahl, U. (1993) *J. Biol. Chem.* **268**(32), 23898-23905

175. Kreuger, J., Prydz, K., Pettersson, R. F., Lindahl, U., and Salmivirta, M. (1999) *Glycobiology* **9**(7), 723-729

176. Chen, J., Avci, F. Y., Munoz, E. M., McDowell, L. M., Chen, M., Pedersen, L. C., Zhang, L., Linhardt, R. J., and Liu, J. (2005) *J. Biol. Chem.* **280**(52), 42817-42825

177. Nakato, H., and Kimata, K. (2002) *Biochim. Biophys. Acta* **1573**(3), 312-318

178. Kamimura, K., Koyama, T., Habuchi, H., Ueda, R., Masu, M., Kimata, K., and Nakato, H. (2006) *J. Cell Biol.* **174**(6), 773-778

179. Parish, C. R. (2006) *Nat Rev Immunol* **6**(9), 633-643

180. Gotte, M. (2003) *FASEB J.* **17**(6), 575-591

181. Giuffre, L. e. a. (1997) *J. Cell Biol.* **136**(945-956)

182. Xie, X., Rivier, A.-S., Zakrzewicz, A., Bernimoulin, M., Zeng, X.-L., Wessel, H. P., Schapira, M., and Spertini, O. (2000) *J. Biol. Chem.* **275**(44), 34818-34825

183. Nelson, R. M., Cecconi, O., Roberts, W. G., Aruffo, A., Linhardt, R. J., and Bevilacqua, M. P. (1993) *Blood* **82**(11), 3253-3258

184. Wang, L., Brown, J. R., Varki, A., and Esko, J. D. (2002) *J. Clin. Invest.* **110**(1), 127-136

185. John, B. L. (2002) *Immunol. Rev.* **186**(1), 19-36

186. Rosen, S. D. (2004) *Ann. Rev. Immunol.* **22**(1), 129-156

187. Wang L, F. M., Sriramarao P, Esko JD. (2005) *Nat. Immunol.* **6**(9), 902-910

188. Parish, C. R. (2005) *Nat. Immunol.* **6**, 861-862

189. Handel, T. M., Johnson, Z., Crown, S. E., Lau, E. K., Sweeney, M., and Proudfoot, A. E. (2005) *Ann. Rev. Biochem.* **74**(1), 385-410

190. Sadir, R., Imberty, A., Baleux, F. o., and Lortat-Jacob, H. (2004) *J. Biol. Chem.* **279**(42), 43854-43860

191. L M Webb, M. U. E., I Clark-Lewis, M Baggiolini, and A Rot. (1993) *Proc Natl Acad Sci U S A.* **90**(15), 7158-7162

192. Revital Shamri, V. G., Jean-Marc Gauguet, Sara Feigelson, Eugenia Manevich, Waldemar Kolanus, Martyn K Robinson, Donald E Staunton, Ulrich H von Andrian & Ronen Alon. (2005) *Nat. Immunol.* **6**, 497-506

193. Chelmicka-Szorc, E., and Arnason, B. G. W. (1972) *Arch Neurol* **27**(2), 153-158

194. Willenborg, D. O., and Parish, C. R. (1988) *J Immunol* **140**(10), 3401-3405

195. Lider, O., Baharav, E., Mekori, Y. A., Miller, T., Naparstek, Y., Vlodavsky, I., and Cohen, I. R. (1989) *J. Clin. Invest.* **83**(3), 752-756

196. Sy MS, S. E., McCluskey R, Greene MI, Rosenberg RD, Benacerraf B. (1983) *Cell Immunol.* **82**(1), 23-32

197. Shieh, M. T., WuDunn, D., Montgomery, R. I., Esko, J. D., and Spear, P. G. (1992) *J. Cell Biol.* **116**(5), 1273-1281

198. Tyagi, M., Rusnati, M., Presta, M., and Giacca, M. (2001) *J. Biol. Chem.* **276**(5), 3254-3261

199. Garson, J. A., Lubach, D., Passas, J., Whitby, K., and Grant, P. R. (1999) *J. Med. Virol.* **57**(3), 238-242

200. Giroglou, T., Florin, L., Schafer, F., Streeck, R. E., and Sapp, M. (2001) *J. Virol.* **75**(3), 1565-1570

201. Shukla, D., and Spear, P. G. (2001) *J. Clin. Invest.* **108**(4), 503-510

202. Spear, D. W. a. P. G. (1989) *J. Virol.* **63**(1), 52-58

203. Liu, J., Shriver, Z., Pope, R. M., Thorp, S. C., Duncan, M. B., Copeland, R. J., Raska, C. S., Yoshida, K., Eisenberg, R. J., Cohen, G., Linhardt, R. J., and Sasisekharan, R. (2002) *J. Biol. Chem.* **277**(36), 33456-33467

204. Herold, B. C., Visalli, R. J., Susmarski, N., Brandt, C. R., and Spear, P. G. (1994) *J Gen Virol* **75**(6), 1211-1222

205. Trybala, E., Bergstrom, T., Svennerholm, B., Jeansson, S., Glorioso, J. C., and Olofsson, S. (1994) *J Gen Virol* **75**(4), 743-752

206. Herold, B. C., WuDunn, D., Soltys, N., and Spear, P. G. (1991) *J. Virol.* **65**(3), 1090-1098

207. Feyzi, E., Trybala, E., Bergstrom, T., Lindahl, U., and Spillmann, D. (1997) *J. Biol. Chem.* **272**(40), 24850-24857

208. Whitbeck, J. C., Peng, C., Lou, H., Xu, R., Willis, S. H., Ponce de Leon, M., Peng, T., Nicola, A. V., Montgomery, R. I., Warner, M. S., Soulika, A. M., Spruce, L. A., Moore, W. T., Lambris, J. D., Spear, P. G., Cohen, G. H., and Eisenberg, R. J. (1997) *J. Virol.* **71**(8), 6083-6093

209. Spear, P. G., Eisenberg, R. J., and Cohen, G. H. (2000) *Virology* **275**(1), 1-8

210. PG, S. (1993) *Viral fusion mechanisms*, CRC Press, Inc., Boca Raton, FL

211. Montgomery, R. I., Warner, M. S., Lum, B. J., and Spear, P. G. (1996) *Cell* **87**(3), 427-436

212. Warner, M. S., Geraghty, R. J., Martinez, W. M., Montgomery, R. I., Whitbeck, J. C., Xu, R., Eisenberg, R. J., Cohen, G. H., and Spear, P. G. (1998) *Virology* **246**(1), 179-189

213. Krummenacher, C., Nicola, A. V., Whitbeck, J. C., Lou, H., Hou, W., Lambris, J. D., Geraghty, R. J., Spear, P. G., Cohen, G. H., and Eisenberg, R. J. (1998) *J. Virol.* **72**(9), 7064-7074

214. Shukla, D., Liu, J., Blaiklock, P., Shworak, N. W., Bai, X., Esko, J. D., Cohen, G. H., Eisenberg, R. J., Rosenberg, R. D., and Spear, P. G. (1999) *Cell* **99**(1), 13-22

215. Anderson, R. A., Feathergill, K., Diao, X., Cooper, M., Kirkpatrick, R., Spear, P., Waller, D. P., Chany, C., Doncel, G. F., Herold, B., and Zaneveld, L. J. (2000) *J Androl* **21**(6), 862-875

216. Raghuraman, A., Tiwari, V., Zhao, Q., Shukla, D., Debnath, A. K., and Desai, U. R. (2007) *Biomacromolecules* **8**(5), 1759-1763

217. Copeland, R., Balasubramaniam, A., Tiwari, V., Zhang, F., Bridges, A., Linhardt, R. J., Shukla, D., and Liu, J. (2008) *Biochemistry* **47**(21), 5774-5783

218. Sasisekharan, R., Shriver, Z., Venkataraman, G., and Narayanasami, U. (2002) *Nat Rev Cancer* **2**(7), 521-528

219. Varki, N. M., and Varki, A. (2002) *Semin Thromb Hemost* **28**(01), 53-66

220. Cosgrove, R. H., Zacharski, L. R., Racine, E., and Andersen, J. C. (2002) *Semin Thromb Hemost* **28**(01), 79-88

221. Kakkar, A. K., and Williamson, R. C. N. (1999) *BMJ* **318**(7198), 1571-1572

222. Tagalakis, V., Blostein, M., Robinson-Cohen, C., and Kahn, S. R. (2007) *Cancer Treat. Rev.* **33**(4), 358-368

223. Borsig, L., Wong, R., Feramisco, J., Nadeau, D. R., Varki, N. M., and Varki, A. (2001) *Proc. Natl. Acad. Sci. USA* **98**(6), 3352-3357

224. Bick, R. L. (2003) *N Engl J Med* **349**(2), 109-111

225. Israel Vlodavsky, Y.F. Michael Elkin. Helena Aingorn, Ruth Atzmon, Rivka Ishai-Michaeli, M.B., Orit Pappo, Tuvia Peretz, Israel Michal, and Pecker, L.S.I. (1999) *Nat. Med.* **5**(793-802)

226. Siro, S., Keisuke, I., and Hiroyuki, O. (2004) *Cancer Science* **95**(7), 553-558

227. Parish, C. R., Freeman, C., Brown, K. J., Francis, D. J., and Cowden, W. B. (1999) *Cancer Res* **59**(14), 3433-3441

228. Uno, F., Fujiwara, T., Takata, Y., Ohtani, S., Katsuda, K., Takaoka, M., Ohkawa, T., Naomoto, Y., Nakajima, M., and Tanaka, N. (2001) *Cancer Res* **61**(21), 7855-7860

229. Eutick, M. (2008) www.progen.com.au

230. Zheng, L., Baumann, U., and Reymond, J.-L. (2004) *Nucl. Acids Res.* **32**(14), e115-

231. Shively JE, C. H. (1976) *Biochemistry* **15**(18), 3932-3942

232. Terpe, K. (2003) *Appl Microbiol Biotechnol* **60**, 523-533

233. Randall, W. C., Hiestand W.A. (1939) *J. Amer. Phys.*, 761-767

234. Fuster, M. M., and Esko, J. D. (2005) *Nat Rev Cancer* **5**(7), 526-542

235. Negishi, M., Pedersen, L. G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y., and Pedersen, L. C. (2001) *Arch. Biochem. Biophys.* **390**(2), 149-157

236. Casu B., L., U. (2001) *Adv Carbohydr Chem Biochem* **57**, 159-206

237. Mikhailov, D., Young, H. C., Linhardt, R. J., and Mayo, K. H. (1999) *J. Biol. Chem.* **274**(36), 25317-25329

238. Shipp, E. L., and Hsieh-Wilson, L. C. (2007) *Chem. Biol.* **14**(2), 195-208

239. Li K, B. H., Liu J. (2010) *J. Biol. Chem.*

240. Muñoz, E., Xu, D., Avci, F., Kemp, M., Liu, J., and Linhardt, R. J. (2006) *Biochem. Biophys. Res. Commun.* **339**(2), 597-602

241. Hagner-McWhirter, A., Li, J.-P., Oscarson, S., and Lindahl, U. (2004) *J. Biol. Chem.* **279**(15), 14631-14638

242. http://www.uniprot.org/uniprot/O94923. In.

243. Rost, B., Yachdav, G., and Liu, J. (2004) *Nucl. Acids Res.* **32**(suppl_2), W321-326

244. Shaya, D., Tocilj, A., Li, Y., Myette, J., Venkataraman, G., Sasisekharan, R., and Cygler, M. (2006) *J. Biol. Chem.* **281**(22), 15525-15535

245. Shriver, Z., Hu, Y., Pojasek, K., and Sasisekharan, R. (1998) *J. Biol. Chem.* **273**(36), 22904-22912

246. Pacheco, B., Malmstrom, A., and Maccarana, M. (2009) *J. Biol. Chem.* **284**(15), 9788-9795

247. Munoz, E., Xu, D., Kemp, M., Zhang, F., Liu, J., and Linhardt, R. J. (2006) *Biochemistry* **45**(16), 5122-5128

248. Sueyoshi, T., Kakuta, Y., Pedersen, L. C., Wall, F. E., Pedersen, L. G., and Negishi, M. (1998) *FEBS Letters* **433**(3), 211-214

249. Pedersen, L. C., Petrotchenko, E., Shevtsov, S., and Negishi, M. (2002) *J. Biol. Chem.* **277**(20), 17928-17932.

250. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) *Nature* **446**(7139), 1030-1037

251. Gama, C. I., Tully, S. E., Sotogaku, N., Clark, P. M., Rawat, M., Vaidehi, N., Goddard, W. A., Nishi, A., and Hsieh-Wilson, L. C. (2006) *Nat Chem Biol* **2**(9), 467-473

252. Copeland R.J., B., A., Tiwari, V., Zhang, F., Bridges, A., Lindhart, R.J., Shukla, D., Liu, J. (2008) *Biochemistry* **47**, 5774-5783

253. Raman, R., Myette, J., Venkataraman, G., Sasisekharan, V., and Sasisekharan, R. (2002) *Biochem. Biophys. Res. Commun.* **290**(4), 1214-1219

254. Beisswanger, R., Corbeil, D., Vannier, C., Thiele, C., Dohrmann, U., Kellner, R., Ashman, K., Niehrs, C., and Huttner, W. B. (1998) *Proc. Natl. Acad. Sci. USA* **95**(19), 11134-11139

255. Niehrs C, B. R., Huttner WB. (1994) *Chem Biol Interact* **92**(1-3), 257-271

256. Stone, M. J., Chuang, S., Hou, X., Shoham, M., and Zhu, J. Z. (2009) *New Biotechnol.* **25**(5), 299-317

257. Moore, K. L. (2003) *J. Biol. Chem.* **278**(27), 24243-24246

258. Fieger, C. B., Sassetti, C. M., and Rosen, S. D. (2003) *J. Biol. Chem.* **278**(30), 27390-27398

259. Dong JF, L. C., López JA. (1994) *Biochemistry* **33**(46), 13946-13953

260. Marchese, P., Murata, M., Mazzucato, M., Pradella, P., De Marco, L., Ware, J., and Ruggeri, Z. M. (1995) *J. Biol. Chem.* **270**(16), 9571-9578

261. Wilkins, P. P., Moore, K. L., McEver, R. P., and Cummings, R. D. (1995) *J. Biol. Chem.* **270**(39), 22677-22680

262. Sako, D., Comess, K. M., Barone, K. M., Camphausen, R. T., Cumming, D. A., and Shaw, G. D. (1995) *Cell* **83**(2), 323-331

263. Pouyani, T., and Seed, B. (1995) *Cell* **83**(2), 333-343

264. Hortin, G. L. (1990) *Blood* **76**(5), 946-952

265. Pittman DD, T. K., Michnick D, Selighsohn U, Kaufman RJ. (1994) *Biochemistry* **33**(22), 6952-6959

266. Leyte, A., van Schijndel, H. B., Niehrs, C., Huttner, W. B., Verbeet, M. P., Mertens, K., and van Mourik, J. A. (1991) *J. Biol. Chem.* **266**(2), 740-746

267. Michnick, D. A., Pittman, D. D., Wise, R. J., and Kaufman, R. J. (1994) *J. Biol. Chem.* **269**(31), 20095-20102

268. Evans, J. P., Watzke, H. H., Ware, J. L., Stafford, D. W., and High, K. A. (1989) *Blood* **74**(1), 207-212

269. White II GC, B. A., Nielsen B. (1997) *Thromb. Haemost.* **78**, 261-265

270. Thompson, A. R. (1986) *Blood* **67**(3), 565-572

271. Farrell DH, M. E., Huang SM, Chung DW, Davie EW. (1991) *Biochemistry* **30**(39), 9414-9420

272. Forbes, E. G., Cronshaw, A. D., MacBeath, J. R. E., and Hulmes, D. J. S. (1994) *FEBS Letters* **351**(3), 433-436

273. Arja, J., Juha, R., Onni, N., and Leila, R. (1986) *Eur. J. Biochem.* **154**(1), 219-224

274. Liu MC, L. F. (1985) *Proc Natl Acad Sci U S A.* **82**(1), 34-37

275. Dieter, J., Annette, H., Keith, K. S., and Wieland, B. H. (1989) *Eur. J. Biochem.* **185**(2), 391-395

276. Hortin, G., Fok, K. F., Toren, P. C., and Strauss, A. W. (1987) *J. Biol. Chem.* **262**(7), 3082-3085

277. Hortin, G., Folz, R., Gordon, J. I., and Strauss, A. W. (1986) *Biochem. Biophys. Res. Commun.* **141**(1), 326-333

278. Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N. P., Gerard, C., Sodroski, J., and Choe, H. (1999) *Cell* **96**(5), 667-676

279. Preobrazhensky, A. A., Dragan, S., Kawano, T., Gavrilin, M. A., Gulina, I. V., Chakravarty, L., and Kolattukudy, P. E. (2000) *J Immunol* **165**(9), 5295-5303

280. Farzan, M., Babcock, G. J., Vasilieva, N., Wright, P. L., Kiprilov, E., Mirzabekov, T., and Choe, H. (2002) *J. Biol Chem.* **277**(33), 29484-29489

281. Fong, A. M., Alam, S. M., Imai, T., Haribabu, B., and Patel, D. D. (2002) *J. Biol. Chem.* **277**(22), 19418-19423

282. Farzan, M., Schnitzler, C. E., Vasilieva, N., Leung, D., Kuhn, J., Gerard, C., Gerard, N. P., and Choe, H. (2001) *J. Exp. Med.* **193**(9), 1059-1066

283. Costagliola S, P. V., Bonomi M, Koch J, Many MC, Smits G, Vassart G. (2002) *EMBO J.* **21**(4), 504-513

284. Bundgaard JR, V. J., Rehfeld JF. (1995) *EMBO J.* **14**(13), 3073-3079

285. Mutt V, J. J. (1968) *Eur. J. Biochem.* **6**(1), 156-162

286. Liu, M. C., Yu, S., Sy, J., Redman, C. M., and Lipmann, F. (1985) *Proc. Natl. Acad. Sci. USA* **82**(21), 7160-7164

287. Weidemann, A., König, G., Bunke, D., Fischer, P., Salbaum, J. M., Masters, C. L., and Beyreuther, K. (1989) *Cell* **57**(1), 115-126

288. Karp, D. R. (1983) *J. Biol. Chem.* **258**(21), 12745-12748

289. Bielinska, M. (1987) *Biochem. Biophys. Res. Commun.* **148**(3), 1446-1452

290. Hsu, Y.-R., Hsu, E. W. J., Katta, V., Brankow, D., Tseng, J., Hu, S., Morris, C. F., Kenney, W. C., and Lu, H. S. (1998) *Protein Exp. Purif.* **12**(2), 189-200

291. Omary MB, d. G. L., Varki NM, Kagnoff MF. (1992) *Mol Immunol* **29**(1), 9-19

292. Baumeister FA, H. V. (1988) *Cell Tissue Res.* **252**(2), 349-358

293. Huttner, W. B. (1988) *Ann. Rev. Physiol.* **50**(1), 363-376

294. Komori, R., Amano, Y., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009) *Proc. Natl. Acad. Sci.* **106**(35), 15067-15072

295. Ouyang, Y.-b., Lane, W. S., and Moore, K. L. (1998) *Proc. Natl. Acad. Sci. USA* **95**(6), 2896-2901

296. Kehoe, J. W., and Bertozzi, C. R. (2000) *Chem. Biol.* **7**(3), R57-R61

297. Borghei, A., Ouyang, Y.-B., Westmuckett, A. D., Marcello, M. R., Landel, C. P., Evans, J. P., and Moore, K. L. (2006) *J. Biol. Chem.* **281**(14), 9423-9431

298. Moore, K. L. (2009) *Proc. Natl. Acad. Sci.* **106**(35), 14741-14742

299. Mishiro, E., Sakakibara, Y., Liu, M.-C., and Suiko, M. (2006) *J Biochem* **140**(5), 731-737

300. Bundgaard, J. R., Vuust, J., and Rehfeld, J. F. (1997) *J. Biol. Chem.* **272**(35), 21700-21705

301. Niehrs, C., Kraft, M., Lee, R. W., and Huttner, W. B. (1990) *J. Biol. Chem.* **265**(15), 8525-8532

302. Kornfeld, R., and Kornfeld, S. (1985) *Ann. Rev. Biochem.* **54**(1), 631-664

303. Ouyang, Y.-B., Crawley, J. T. B., Aston, C. E., and Moore, K. L. (2002) *J. Biol. Chem.* **277**(26), 23781-23787

304. Westmuckett, A. D., Hoffhines, A. J., Borghei, A., and Moore, K. L. (2008) *Gen. Comp. Endocrin.* **156**(1), 145-153

305. Ouyang, Y.-B., and Moore, K. L. (1998) *J. Biol. Chem.* **273**(38), 24770-24774

306. Colvin, R. A., Campanella, G. S. V., Manice, L. A., and Luster, A. D. (2006) *Mol. Cell. Biol.* **26**(15), 5838-5849

307. Gao, J.-m., Xiang, R.-l., Jiang, L., Li, W.-h., Feng, Q.-p., Guo, Z.-j., Sun, Q., Zeng, Z.-p., and Fang, F.-d. (2009) *Acta Pharmacol Sin* **30**(2), 193-201

308. Moore, J. P., Trkola, A., and Dragic, T. (1997) *Curr. Opin. Immunol.* **9**(4), 551-562

309. Dragic, T. e. a. (1996) *Nature* **381**, 667-673

310. Feng, Y. e. a. (1996) *Science* **272**, 872-877

311. Alkhatib, G. e. a. (1996) *Science* **272**, 1955-1958

312. Farzan, M., Choe, H., Vaca, L., Martin, K., Sun, Y., Desjardins, E., Ruffing, N., Wu, L., Wyatt, R., Gerard, N., Gerard, C., and Sodroski, J. (1998) *J. Virol.* **72**(2), 1160-1164

313. Cormier, E. G., Persuh, M., Thompson, D. A. D., Lin, S. W., Sakmar, T. P., Olson, W. C., and Dragic, T. (2000) *Proc. Natl. Acad. Sci. USA* **97**(11), 5762-5767

314. Hyeryun, C., Michael, J. M., Christopher, M. O., Paulette, L. W., Natalya, V., Wenhui, L., Agam, P. S., Rushdi, S., Chetan, E. C., and Michael, F. (2005) *Mol. Microbiol.* **55**(5), 1413-1422

315. Pittman, D. D., Wang, J. H., and Kaufman, R. J. (2002) *Biochemistry* **31**(13), 3315-3325

316. Pittman, D. D., Tomkinson, K. N., Michnick, D., Selighsohn, U., and Kaufman, R. J. (2002) *Biochemistry* **33**(22), 6952-6959

317. White, I. I. G. C., Pickens, E. M., Liles, D. K., and Roberts, H. R. (1998) *Transfus. Sci.* **19**(2), 177-189

318. Nelsestuen, G. L., Zytkovicz, T. H., and Howard, J. B. (1974) *J. Biol. Chem.* **249**(19), 6347-6350

319. Fernlund, P., and Stenflo, J. (1983) *J. Biol. Chem.* **258**(20), 12509-12512

320. Huttner, W. B. (1982) *Nature* **299**, 273-276

321. Davie, K. K. a. E. W. (1982) *Proc Natl Acad Sci U S A.* **79**(21), 6461–6464

322. Bond M.D. Huberty, M. C., Jankowski, M.A., Vath, J.E., Strang, A-M, and Scoble, H.A. (1997) *Blood* **84**(10), 531a (In Supplement 531)

323. Lin, W. H., Larsen, K., Hortin, G. L., and Roth, J. A. (1992) *J. Biol. Chem.* **267**(5), 2876-2879

324. Hsu, W., Rosenquist, G. L., Ansari, A. A., and Gershwin, M. E. (2005) *Autoimmun. Rev.* **4**(7), 429-435

325. Danan, L. M., Yu, Z., Hoffhines, A. J., Moore, K. L., and Leary, J. A. (2008) *J. Am. Soc. Mass Spectrom.* **19**(10), 1459-1466.