Virginia Commonwealth University

VCU Scholars Compass

2008

Designing Non-saccharide Heparin/Heparan Sulfate Mimics

Arjun Raghuraman
Virginia Commonwealth University

www.manaraa.com

DESIGNING NON-SACCHARIDE HEPARIN/HEPARAN SULFATE MIMICS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

ARJUN RAGHURAMAN
BS in Pharmacy, SRM Institute of Science and Technology, India, 2001

Director: UMESH R. DESAI
PROFESSOR, DEPARTMENT OF MEDICINAL CHEMISTRY

Virginia Commonwealth University
Richmond, Virginia
April, 200

Acknowledgement

First and foremost, I would like to thank my advisor Umesh R. Desai for being a great mentor. He has supported me at every stage of the graduate program and has always been available for discussions. Through his animated anecdotes and analogies, he has guided me along my journey while at the same time provided me with ample room to carve my own path. He has always encouraged me to write papers and research grants, and attend scientific meetings and conferences. I feel that I have received a well-rounded training and for this I am most grateful.

I would also like to thank my committee members Drs. Richard Westkaemper, Glen Kellogg, Vladimir Sidorov and Tonie Wright. This committee has greatly contributed to my scientific development by providing a very challenging Defense of Independent Research Proposal experience. The courses "Advanced Medicinal Chemistry III" and "Advanced Spectroscopy" taught by Drs. Westkaemper and Sidorov, respectively, had a profound intellectual impact on me. I would like to thank Dr. Kellogg for guiding me through my "exit" phase by critiquing cover letters for postdoctoral fellowships and providing advice on applications. I also thank Dr. Wright for reviewing my predoctoral fellowship grant.

I have had the priviledge of working with a number of interesting and talented individuals during my graduate career. I thank the Desai group postdoctoral scholars - Dr.

Monien, Dr. Gunnarsson, Dr. Riaz and Dr.Liang for helping me out with my work at some point of time. I also like to thank the former and current graduate students of our group – Chandravel Krishnasamy, Junaid "the king of man": Afridi, Mohammed Rahman, Manadakini Dantuluri, Jay "the socialist" Thakkar, Brian "the king" Henry, and Tim "the devoted" King for making my graduate experience a memorable one. I would particularly like to thank Chandravel Krishnasamy, Junaid Afridi and Mohammed Rahman for helping me out when I first came to VCU.

Outside of the Desai group, I would like to thank Dr. Phil Mosier for teaching me some "Jedi" tricks in the form of Unix and vi commands. I also thank Drs. Michael Hindle and Yun Qu for providing assistance with mass spectrometry and NMR spectroscopy, respectively. I thank Dr. Lemont Kier for valuable discussions on modeling, in the broadest sense of the word, and for supporting me through my graduate career.

Finally, I would like to thank my parents –Geetha and T.V. Raghuraman for their love and prayers. I also thank my brother and sister-in-law - Vivek and Cathy Ayer, and my uncle Dr. Raghavan Ayer for making sure that I had fun during some often difficult times. I thank Ms. Divya Srinivasan for her unwavering support. This Dissertation would not have been possible without the support of all my friends whom I whole-heartedly thank.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AT: Antithrombin, DS: Dermatan sulfate, DEFGH: Antithrombin-binding heparin pentasaccharide sequence; also known as H5, DEF: trisaccharide corresponding to the non-reducing end of DEFGH, DMA: Dimethylacetamide, DMF: Dimethylformamide, EGF: Epidermal growth factor, EHBS: Extended heparin-binding site, FGF-2: Fibroblast growth factor-2, FGF-2R: Fibroblast growth factor-2 receptor, FX: Fondaparinux, GAG: Glycosaminoglycan, GOLD: Genetic optimization for ligand docking, gB: glycoprotein B, gC: glycoprotein C, gD: glycoprotein D, gp120: glycoprotein 120, GlcAp: Glucuronic acid, GlcNp: D-Glucosamine, HS: Heparan sulfate, HCII: Heparin Cofactor II, HSV: Herpes simplex virus, HCV: Hepatitis C virus, HIV: Human immunodeficiency virus, H5: Antithrombin-binding heparin pentasaccharide sequence; also known as DEFGH, HINT: Hydropathic interaction analyses, HIT: Heparin-induced thrombocytopenia, HL-GAG: Heparan-like glycosaminoglycan; IdoAp: Iduronic acid, LMWH: Low molecular weight heparin, NA domain: N-acetyl domain, NS domain: N-sulfate domain, 3-OST: 3-O-sulfotransferase, PG: Proteoglycan, PBS: Pentasaccharide-binding site, PCI: Protein- C inhibitor, RMSD: Root mean squared deviation, RP-HPLC: Reversed phase high performance liquid chromatography, RCL: Reactive center loop, TNS: 2-p-toluidinonaphthalene-6-sulfonate, TF: Tissue factor, TNF-α: Tissue necrosis factor alpha, UAp: Uronic acid, VEGF: Vascular endothelial growth factor.

Abstract

DESIGNING NON-SACCHARIDE GLYCOSAMINOGLYCAN MIMICS

By Arjun Raghuraman, PhD

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophyat Virginia Commonwealth University.

Virginia Commonwealth University, 2008

Major Director: Umesh R. Desai
Professor, Department of Medicinal Chemistry

Glycosaminoglycans (GAGs) are complex biopolymers that play important roles in inflammation, coagulation, angiogenesis, cell adhesion and viral invasion by interacting with several different proteins.[1,2] Structurally, GAGs are built up of several different sulfated disaccharide units.[3] Specific GAG sequences that uniquely recognize their cognate proteins exist. Such specificity typically arises from the binding of unique sulfation patterns on the linear GAG chain to highly electropositive protein domains. Thus, these highly charged, sulfated biopolymers potentially represent a new class of therapeutics. Yet,

xvi

the major stumbling block to the development to these agents is their extremely complicated and tedious chemical synthesis. We hypothesized that replacing the saccharide skeleton with an equivalent non-saccharide and readily synthesized organic skeleton would usher in an era of new, GAG-based therapeutics. This challenge has been addressed on two fronts, computational design and chemical synthesis, by focusing on the heparin pentasaccharide-antithrombin system that represents an exhaustively studied model GAG-protein system. With respect to chemical synthesis, a microwave-based synthetic procedure that can rapidly introduce multiple sulfate groups on a poly-hydroxyl substrate within minutes was developed.[4] Using this method, the synthesis of a previously designed activator (IAS5), which otherwise proved to be problematic, was successfully completed. Biochemical screening of IAS5 and its analogs revealed that these molecules could activate antithrombin up to 30-fold in comparison to the 300-fold activation by the heparin pentasaccharide. In an effort to develop more potent antithrombin activators, a new method to predict high affinity GAG sequences for a given GAG-binding protein based on combinatorial virtual-library screening was developed.[5] This combinatorial virtual-library screening method was applied to a library of 24,576 non-saccharide, sulfated molecules that were created using the structure of IAS5 as a template. Thirty seven 'hits' that had common structural features were identified from this study. Interestingly, all these 'hits' bind to antithrombin similarly and orient the 4 negative charges identical to the corresponding groups in the heparin pentasaccharide. The synthesis of selected targets is currently in progress and several synthetic steps have already been optimized.

INTRODUCTION

Heparin and heparan sulfate belong to a class of linear, sulfated polysaccharides known as glycosaminoglycans (GAGs). These complex and heterogeneous macromolecules play fundamental roles in a plethora of processes such as growth factor signaling, hemostasis, morphogenesis, inflammation, enzyme regulation and viral invasion.[2]

First introduced in 1916,[6] heparin is the prototype of a class of important anticoagulant drugs used in the clinic today.[7] The prevention of postoperative thrombosis and the treatment of acute venous thrombosis are among heparin's established uses. While heparan sulfate (HS) is commonly thought to be similar to heparin, it is a distinct GAG with significant differences. HS is less sulfated, more heterogeneous and structurally diverse than heparin, and is tagged to core proteins on cell surfaces.[8] Virtually every cell type in metazoan organisms contains HS. In contrast, heparin is found exclusively in mast cells and is not associated with proteins. Thus, HS is ideally poised to mediate cell-cell and cell-matrix interactions.

Investigations into the biosynthesis of these GAGs in the recent past have significantly broadened our understanding of their physiological roles. The basic lesson is

1

that the biosynthesis results in incomplete structural modifications and consequently
innumerable saccharide sequences that modulate the activity of several different proteins in
our body. A partial list of heparin and HS-binding proteins may be found in Table 1.

Table 1. Partial list of GAG-binding proteins (Adapted from Nugent, M.A. et al.,
Chemistry & Biology of Heparin and Heparan Sulfate, 2005, Eds: Garg, Linhardt, Hales,
Ch 19, pg 537)

Growth Factors	ECM Components	Inflammation/Angiogenesis
FGFs	Fibronectin	Interferon γ
TGF β 1&2	Laminins	TNF α
VEGF	Vitronectin	L-selectin
PDGF	Neutrophil elastase	P-selectin

EGF
  Amphiregulin
  Heparin binding EGF
  Betacellulin
  Neuregulin
	Cathepsin	Endostatin

IGF-II	Antithrombin	Angiostatin
Activin	Thrombin	RANTES
TGF β binding protein	Tissue plasminogen activator	Neutrophil activating factor
HGF	Plasminogen activator inhibitor	GM-CSF

A selection of heparin/HS-binding proteins. The abbreviations are: FGF, fribroblast growth
factor; TGF, transforming growth factor; VEGF, vascular endothelial growth factor;
PDGF, platlet-derived growth factor; EGF, epidermal growth factor; IGF, insulin-like
growth factor; TNF, tumor necrosis factor; GM-CSF, granulocyte-macrophage colony
stimulating factor.

1.1. Physiological Roles of Heparin and Heparan Sulfate

1.1.1. Inhibition of Coagulation and Thrombosis

Coagulation is the first line of defense against trauma of the vascular system in humans.[9] Yet, aberrant clotting is the most common cause of death in the industrialized world.[10] The serine proteinase inhibitors (serpins), antithrombin (AT)[11] and heparin cofactor II (HCII),[12] are circulating anticoagulants that serve to prevent thrombosis under physiological conditions. The circulating concentrations of both these inhibitors is significant (1.2 and 2.3 μM for HCII and AT, respectively).[13,14]

AT is a major regulator of the clotting cascade in humans. In fact, AT is essential for survival as homozygous null mutant mice of AT die *in utero*.[15] In addition, numerous cases of AT deficiency, both congenital and acquired, lead to enhanced risk of thrombosis.[16-19] On the other hand, the physiological relevance of HCII remains unclear at the present time. Although HCII deficiency is not a significant risk factor for thrombosis, recent studies suggest that HCII plays an important role during extra-vascular injury and in preventing arterial thrombosis.[20-22] While AT functions as an anticoagulant primarily by inhibiting the procoagulant proteases thrombin, factor Xa and factor IXa, HCII specifically inhibits thrombin. Additionally, HCII is able to inhibit clot-bound thrombin while AT cannot.[23]

AT and HCII can inhibit their target proteases at physiologically relevant rates only in the presence of GAGs.[24,25] The binding of GAG chains to these serpins induces a major conformational change resulting in enhanced recognition of target proteases. This process

is called 'activation'. In addition to conformational activation, GAG chains can bring about

rate accelerations by serving as bridging templates for the inhibitor and protease. While

only heparin and HS can accelerate the rate of AT inhibition of coagulation proteases,

other GAGs such as dematan sulfate (DS) are also effective for HCII (Figure 1).

The conformational activation of AT can be brought about by as little as a 5-

residue heparin/HS sequence that is commonly referred to as sequence-specific heparin
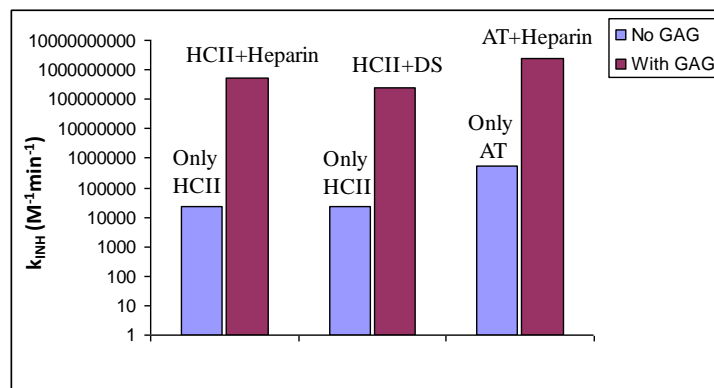
Figure 1. GAG activation of AT and HCII. k_{INH} is the second order rate constant for the serpin-thrombin reaction

pentasaccharide (H5, see Figure 7A for structure).[26,27] The binding of H5 to AT causes

some 300-600-fold acceleration in the inhibition of factors IXa and Xa (Figure 2).[28-30] In

contrast, thrombin inhibition by AT is dependent on GAG chains that contain at least 18

monosaccharides.[31] Although the bridging mechanism is the overriding factor in this case,

physiologically relevant inhibition of thrombin by AT is achieved only with heparin chains

that contain the H5 sequence.

HCII binds heparin through an induced-fit mechanism similar to that of AT.[32] Yet, HCII binds heparin chains 1000-fold more weakly (K_D = 26 μM)[32] than does AT (K_D = 20 nM) to heparin chains containing the H5 sequence.[33] This result is surprising because AT and HCII share significant structural similarities.[34] Heparin chains containing as little as 10 monosaccharides can activate HCII and produce approximately 1000-fold acceleration of thrombin inhibition.[32] Thus, the major mechanism of HCII-rate acceleration appears to be conformational activation as is also suggested by mutagenesis studies of thrombin exosite
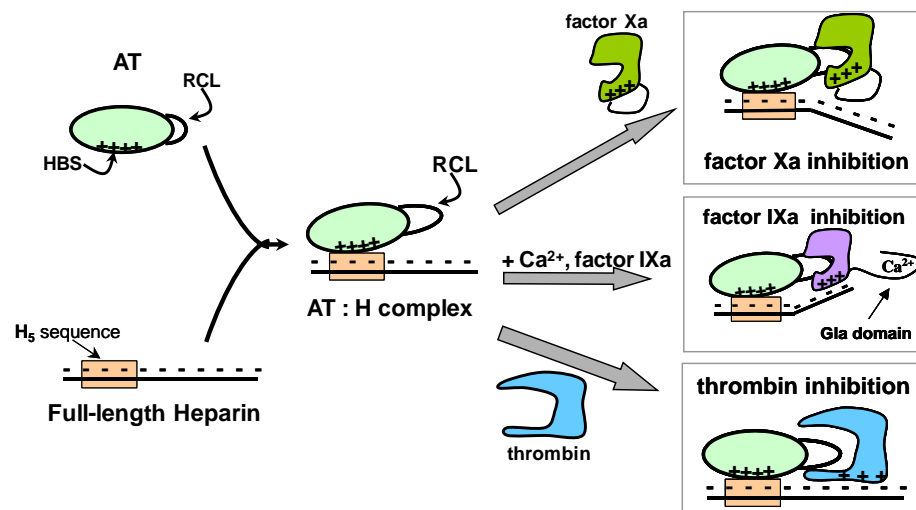
Figure 2. Two major mechanisms of heparin activation of AT inhibition of factor Xa, factor IXa and thrombin - conformational activation and bridging mechanism. AT:H = antithrombin–heparin complex; H_5 = high-affinity pentasaccharide sequence in heparin; RCL = reactive center loop; '+++' = exosite on enzyme; HBS = heparin-binding site (Adapted from Desai, U. R. Med. Res. Rev. 2004, 24, 151-181.)

mutants.[35] While recent studies suggest that the heparin-HCII interaction follows a non-specific binding model,[32] it is important to point out that heparin may not be the physiological activator of HCII, but only serves as a binding model for HS that lines the vascular walls and extravascular spaces.

The acceleration of serpin activity by GAGs is not just limited to AT and HCII but appears to be a general feature of coagulation-regulation by GAGs. Of the 35 serpins that have been identified in the human genome, five are known to bind heparin and HS. These include protease nexin-1[36], protein C inhibitor (PCI)[37,38] and plasminogen activator inhibitor-1[39], in addition to AT and HCII. The bridging mechanism of GAGs appears to be the primary mechanism of rate acceleration in these cases.

1.1.2. Cell Growth Control

Since the original observation that heparin is a potent inhibitor of vascular smooth muscle cell growth in 1977,[40] several studies have shown that heparin and HS interact with a variety of growth factors and cytokines to promote and inhibit cell proliferation during normal tissue development as well as in disease states.

Growth factors are a class of 23 relatively small soluble proteins that produce a wide range of cellular responses such as proliferation, migration and differentiation through their action on specific cell surface receptors.[41] During the development of methods for the isolation and purification of growth factors, many of them were found to bind to heparin-agarose columns.[42] Since this original observation, it is now recognized

that heparin and HS play important roles in cell growth regulation by binding to growth factors.

The prototypic HS-binding growth factor is fibroblast growth factor 2 (FGF 2).[43,44] HS participates in forming a ternary complex by interacting with both FGF II (K_D = 39 nM) and its receptor (K_D = 3.2 μM).[45,46] Kinetic analyses have revealed that HS stabilizes this interaction primarily by lowering the dissociation rate without impacting the association rate. It has also been suggested that GAG chains assist in inducing FGF-2 receptor dimerization and autophosphorylation although the physical orientation of the ligand (FGF-2), receptor and GAG remain controversial despite several high-resolution crystal structures.[47-54] Similarly, HS appears to enhance the over-all affinity of vascular endothelial growth factor (VEGF)[55] and epidermal growth factor (EGF)[56] to their cognate receptors. It is interesting to note that endothelial cells treated with heparanases (heparin/HS degrading enzymes) show significant reduction in the binding to VEGF121 even though this isoform does not contain a HS-binding domain. In this case mechanisms involving direct cell surface interactions between HS and the VEGF121receptor appear to be predominant.[57] Thus, it is likely that a bridging mechanism is not the only mechanism utilized by HS in modulating growth factor-growth factor receptor interactions.

The ability of HS to control growth factor- receptor interactions has also been shown to alter receptor signaling, although these two events do not correlate completely.[58-63] The effect of HS on growth factor receptor signaling is complicated by the fact that HS can directly modulate cell growth through mechanisms that are independent of growth

factor receptors. These mechanisms are not dependent on protein ligands and involve the direct interaction of heparin/HS with cell surface "heparin" receptors.[64,65]

Since heparin and HS are intimately involved with cell growth control, it is not surprising that these molecules are involved in cancer.[66] Indeed, heparin and HS are implicated in nearly all stages of cancer such as tumorigenesis,[67,68] angiogenesis,[69,70] and tumor invasion and metastasis.[71,72]

1.1.3. Role of Heparin and HS in Inflammation

There are several lines of evidence that point towards links between thrombosis and inflammation in vascular, cardio-vascular and inflammatory diseases.[73] For example, pro-inflammatory stimuli resulting from interleukins or exposure to E. coli endotoxin increase the levels of tissue necrosis factor-α (TNF-α) and other cytokines, which in turn leads to the activation of leukocytes. Activated leukocytes shed soluble L-selectin from their membranes, which leads to the generation of tissue factor (TF). TF initiates and amplifies a hypercoagulable state resulting in thrombin generation and platelet activation. Hence, molecules that regulate coagulation may also regulate inflammation.

A large body of evidence supports the concept that heparin/HS has anti-inflammatory actions.[74-76] Apart from the modulation of the pathophysiological effects of endotoxin and TNF-α,[77,78] heparin has been shown to suppress selected neutrophil functions such as superoxide generation and chemotaxis *in vitro*.[79,80] The binding of heparin and HS to adhesion molecules expressed on the endothelial and/or leukocyte cell surfaces appears to be the mechanism of anti-inflammatory activity (see Table 1). Indeed,

several clinical studies have suggested that heparin and HS may be therapeutic in the management of ulcerative colitis and Crohn's disease.[81,82] Interestingly, no hemorrhagic complications were observed during these studies. Since the anticoagulant effects of heparin are critically dependent on the H5 sequence, it is likely that heparin's anti-inflammatory effects are distinct from its anticoagulant effects.

Heparin-protein interactions also regulate the complement system. Comprised of about 25 proteins, the complement system is a major defense system that facilitates phagocytosis and bacterial cell lysis.[83] Since the original report by Ecker and Gross in 1929,[84] several reports have shown that heparin/HS regulates multiple steps in the complement system by binding to complement proteins.[85-88] Yet, most of these studies were qualitative and lacked kinetic and thermodynamic data. More recently, surface plasmon resonance (SPR) has been exploited to measure kinetic and thermodynamic constants for the interaction of heparin with several different complement proteins.[89] These studies revealed that heparin binds most complement proteins with an affinity between 10 and 400 nM. By factoring in the concentration of different complement proteins in the plasma, these studies may provide some insight into the design of therapeutic approaches to regulate complement activation.

1.1.4. Heparin and HS mediate Viral Entry into Cells

The ubiquitously expressed eukaryotic cell surface HS serves as binding sites to several different viruses, most notably Herpes Simplex Virus (HSV) 1 and 2,[90] Hepatitis C Virus (HCV)[91] and Human Immunodeficiency Virus 1 (HIV-1).[92,93] These viruses contain

several glycoproteins integrated into a lipid bilayer-based envelope. Typically, viral entry takes place by the fusion of the viral envelope and the eukaryotic cell membrane. By recognizing viral glycoproteins, HS mediates the first step in the entry of viruses into cells. Although this step can be thought of as a non-specific anchoring event, recent evidence suggests that HS plays a role that goes well beyond a non-specific paradigm in certain cases.

Several lines of evidence support the requirement of HS for the entry of HSV-1 into cells. Cells treated with heparanases (HS cleaving enzymes) or that are genetically altered to prevent HS biosynthesis show reduced capacity to bind the virus.[94,95] Soluble heparin or HS inhibit HSV-1 infection,[94,96] while soluble forms of the viral glycoproteins gB and gC of HSV-1 bind HS.[97-99] More interestingly, a third glycoprotein gD has been shown to bind to only a subset of HS sequences ($K_D = 2 \mu M$) suggesting specificity in recognition.[100,101] Given that the binding of gD to cell surface receptors triggers the crucial step of fusion, this result can have therapeutic implications. It is important to point out that although HS is not absolutely required for viral invasion, it greatly increases the efficiency of infection.

Although similar evidence exists in the case of HSV-2, it appears that HSV-2 recognizes different structural features in HS,[102,103] and differs in the relative importance of HS-binding glycoproteins for the initial step of binding. Other viruses belonging to the herpes super family of viruses also require HS for the initial binding event. These include human cytomegalovirus (CMV), human herpes virus (HHV) and varicella zoster virus (VZV).[90] Considering that the lysine-rich HS –binding domains of glycoproteins gB and

gC are conserved across these viruses, this result is not surprising. Similar conservation of positively charged residues in the N-terminus of hepatitis-C virus' envelope glycoprotein E2 implicates the involvement of HS.[91] Indeed, heparin binds to E2 with an affinity of 5.2 nM, while deletion of E2 hypervariable region-1 significantly reduced interaction with HS, suggesting that the collection of positively charged residues present in this region forms the HS-binding site.

The situation for HIV-1 is particularly interesting because unlike the other viruses discussed above, HS binds gp120 of the HIV-1 virus at a late stage.[93] Specifically, CD4, the primary receptor for HIV-1, induces a conformational change in gp120 that dramatically increases its affinity for HS. The domain formed by this conformational change (called the CD4-induced epitope) together with the V3 loop in gp120 form the HS binding sites. Importantly, this binding site is shared by other viral co-receptors suggesting potential therapeutic applications for HS mimetics. Other than gp120, HS has been shown to bind the Tat protein, a potent transcriptional activator of HIV-1, and mediate its internalization.[92] Notably, essentially all Tat-mediated transcriptional activation is dependent on HS in a cell line that specifically expresses the HS proteoglycan perlecan.

1.2. Chemical Structure of Heparin and HS

Both heparin and HS are linear, unbranched polysaccharides composed of disaccharide units consisting of a hexuronic acid 1,4-linked to a *D*-glucosamine unit (**Figure 3**A).[2] The hexuronic acid can be *D*-glucuronic acid (GlcA*p*) or its C5 epimer, L-iduronic acid (IdoA*p*). While GlcA*p* units are largely unmodified, IdoA*p* units are
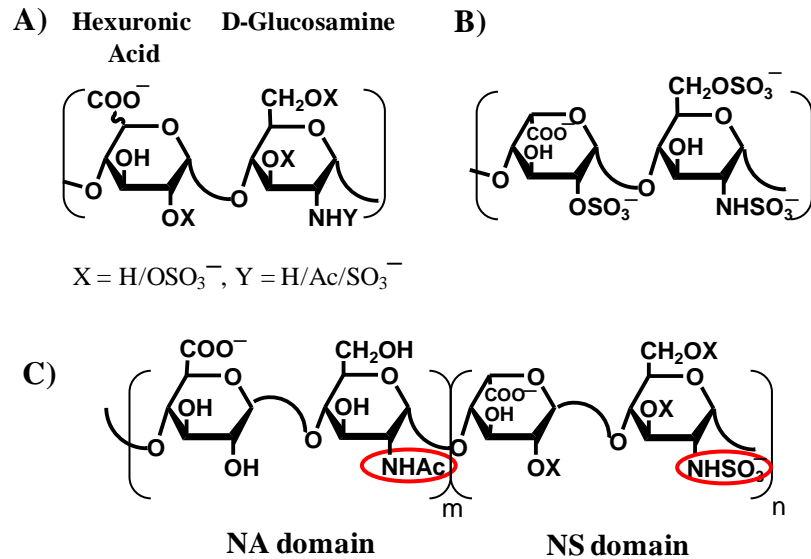
Figure 3. A) Variable disaccharide of heparin and HS. B) Preponderant disaccharide of heparin. C) Domain architecture of HS

frequently modified by 2-*O*-sulfation. *D*-glucosamine residues (GlcN*p*) may be *N*-acetylated or *N*-sulfated, while *N*-sulfated residues may be further modified by 3 and/or 6-*O*-sulfation. These distinct possibilities and combinations have resulted in the identification of as many as 23 different disaccharide building blocks to date (Figure 4).[104] Thus, heparin and HS contain a staggering number of differentially sulfated saccharide sequences.

Heparin has a molecular weight ranging from 5 to 40 kDa with an average of about 15 kDa.[105] The majority of heparin's uronic acid (> 70 %) residues are *L*-iduronic acid.[106,107] A prototypical heparin disaccharide contains three sulfate groups rendering heparin the most acidic polymer in our body (Figure 3B). HS chains tend to be longer than

heparin and vary from 5 to 50 kDa with an average molecular weight of 30 kDa. HS is considerably enriched in unsulfated *N*-acetyl GlcN*p* and GlcA*p* disaccharides. These disaccharides are present contiguously to form *N*-acetyl domains (NA domains, Figure 3C).[108] Relatively shorter segments of sulfated disaccharides containing IdoA*p* and *N*-sulfated GlcN*p* derivatives (NS domains) are found between two NA domains. Interestingly, HS chains also contain some sections of mixed NA/NS domains with moderate degrees of sulfation. It is hypothesized that these mixed NA/NS domains constitute specific protein-binding motifs.[1] It is important to note that the domain architecture of HS is absent in heparin, which may be considered as an extended NS domain of HS.

1.2.1. Conformation of Heparin and HS

Given the numerous biological roles of GAGs, several different research groups have attempted to understand the conformation of these molecules (full length polymers as well as oligosaccharides) through NMR studies.[109-114] In addition, several GAG-protein crystal structures that have emerged in the past decade have helped corroborate conclusions that were derived from solution-based NMR studies.[50,52,53,115-122] Some groups have studied the conformation of GAGs by X-ray diffraction of polysaccharide films.[123,124] Collectively, these studies have pointed out that, unlike proteins, GAGs do not fold into globular structures but remain linear in solution. Specifically, the GAG backbone retains a helical conformation wherein the exact helical parameters (*n*, the number of disaccharides

GlcAp disaccharides   **IdoAp disaccharides**
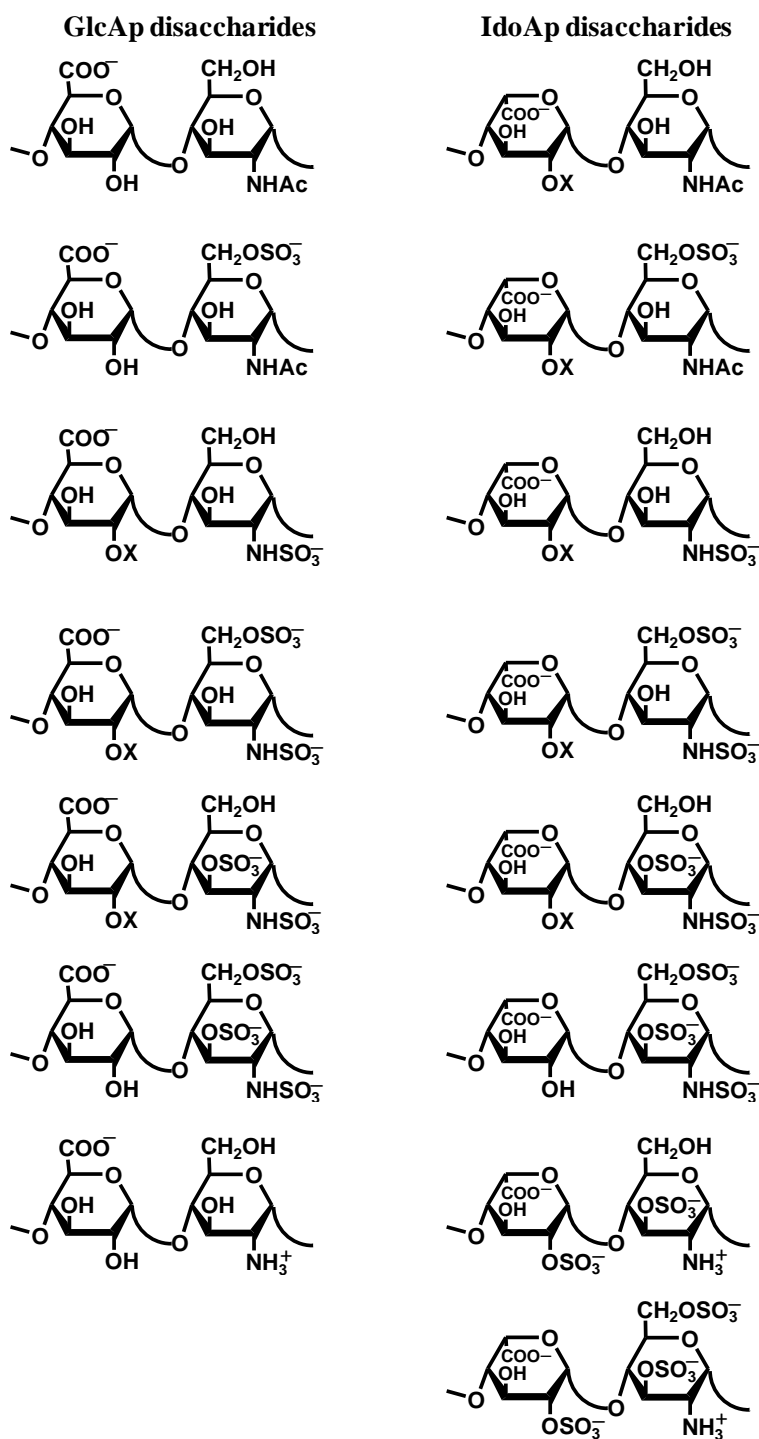
Figure 4. Disaccharide building blocks found in heparin/HS (X=H/SO$_3$)

per turn and *h*, the axial rise per disaccharide) depend on the nature of cationic counterions. For the sodium form of both heparin and HS, n=2 and h=8.4 Å.

Given the helical symmetry, the precise topology of the GAG chain is determined by the glycosidic bond torsions and pyranose ring conformations. For heparin/HS, GlcN*p* and GlcA*p* residues adopt a relatively rigid $^4C_1$ chair conformation, while IdoA*p* residues are more flexible and have the ability to populate multiple low-energy forms such as the $^1C_4$ and $^4C_1$ chair forms and the ^2S_O skew-boat form (Figure 5).[125] While internal IdoA*p*±2S residues reside predominantly in the $^1C_4$ chair (60 %) and ^2S_O skew-boat (40 %) forms, the equilibrium is displaced towards the ^2S_O form when an IdoAp2S residue is preceded by a 3-*O*-sulfated aminosugar.[114,125,126] For terminal unsulfated IdoAp residues, the $^4C_1$ chair form is also possible.

$$^4C_1 \qquad ^1C_4 \qquad ^OS_2 \qquad ^2S_O$$
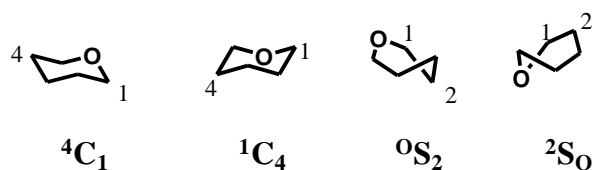
Figure 5. Major IdoA*p* conformations that exist in GAGs

1.3. Biosynthesis of GAGs

HS chains are assembled while attached to a proteoglycan (PG) core protein. Three major families of core PG core proteins have been characterized: the syndecans,[127] the glypicans,[128] and the basement membrane PG perlecan.[129] While the syndecans are

membrane-spanning proteins, glypicans are anchored to the membrane via a phospholipid tether. Although these different core proteins are expressed in a cell-type specific manner, the structure of HS does not appear to correlate with the core proteins but rather on the cell type of origin.[130] Heparin is also synthesized on a core protein called serglycin in mastoctyes. However, after biosynthesis, heparin is shed from the core protein and remains associated with mast cell granules.

Heparin and HS are assembled via a similar pathway by over 30 different enzymes.[104] Chain initiation occurs at the Golgi apparatus by the action of 4 different glycosyl transferases. The resulting tetrasaccharide attached to a specific serine residue on the core protein through an *O*-glycosidic bond, serves to prime chain elongation. Subsequently, *N*-acetyl-*D*-GlcNp residues and *D*-GlcAp residues are added in an alternating fashion to the non-reducing end of the nascent GAG chain. Two glycosyl transferases (EXT1 and EXT2) that form hetero-oligomeric complexes in the Golgi apparatus are responsible for the disaccharide addition. When these disaccharides remain unmodified, they constitute the NA domains of HS. As the polysaccharide chain forms, it is simultaneously modified by at least four different families of sulfotransferases and one epimerase. *N*-deacetylation and *N*-sulfation is the first of these modifications and is carried out by a multi-functional *N*-deacetylase/*N*-sulfotransferase. A C5 epimerase and different O-sulfotransferases perform subsequent modifications to generate different saccharide sequences. Interestingly for HS, these modifications remain localized forming short stretches of NS domains.

The sequence diversity of HS is governed by the controlled cell-type dependent expression of the biosynthetic enzymes and the presence of distinct enzyme isoforms with unique substrate specificities. Interestingly, the type of disaccharides and the relative content of NA, NS and NA/NS domains found in HS appear to be a stable characteristic of the cell type of origin.[131-136] Not surprisingly, while the theoretical number of disaccharide units considering all combination of structural modifications is 48, only 23 have been identified (Figure 4).[104] The sulfation states of HS are also dynamically regulated in embryonic cells during development and in adult normal and cancer cells that respond to growth factor signals.[137-140] In addition, once HS chains have been synthesized, their sulfation states are subject to considerable degree of post-biosynthetic remodeling by sulfatase enzymes. Recently, a novel family of cell surface 6-O-endosulfatases has been discovered.[141] These enzymes are distinguished in two respects: 1) they have a unique "endolytic" mode in contrast to the "exolytic" mode of action of other HS sulfatases,[142] and 2) they play a key role in signaling pathways.[141] Thus, HS sequence diversity is also governed by the controlled expression of sulfatases in cells.

1.4. The Sulfation Code and Specificity in the Interaction of GAG with Proteins

From the discussion above, it is apparent that GAGs can modulate protein activity in two fundamentally different ways. GAGs may serve as non-specific templates to proteins. Such templates may serve to a) reduce entropic costs of biological reactions, (see section 1.1.1) b) serve as a repository for extra-cellular proteins like growth factors,[143,144] and c) help maintain protein gradients across cells.[145] Indeed, GAGs have been shown to

maintain morphogen gradients across cells and tissues that are essential for developmental processes.[145] Maintaining such gradients would involve graded affinities between GAG sequences and protein. In such paradigms, it is the non-specificity in GAG-protein interactions that results in biologically significant functions.

GAGs may also modulate protein function through specific high-affinity interactions. Such specific interactions are typically dependent on the presence of a unique set of modifications localized to a region of the GAG chain. In such a specific paradigm, it appears that the distribution of sulfate groups on a GAG chain dictates which protein it would bind. A limited number of specific GAG sequences have been elucidated to date.

The most thoroughly researched specific GAG sequence is the antithrombin-binding H5 sequence (Figure 3A). This sequence was discovered in the early 1980s by the Lindahl and Rosenberg groups and found to be essential for heparin's anticoagulant activity.[146,147] The presence of the 3-O-sulfate in ring F and the D-glucuronic acid residue are striking features given that most of heparin consists of an IdoAp-based trisulfated disaccharide (Figure 3B). A second interesting specific sequence appears to be the HSV-1 gD-binding sequence (Figure 6B). This sequence was reported in 2002 and contains a rare GlcNH$_3$3S residue.[148] The precise configuration of the hexuronic acid rings and the minimal size required for binding remain unknown at the present time. It is worth mentioning that the 3-O-sulfotransferase isoform (3-OST3) that installs the 3-O-sulfate group in this sequence is different from the one that installs the 3-O-sulfate group in the H5 sequence (3-OST1).[90]
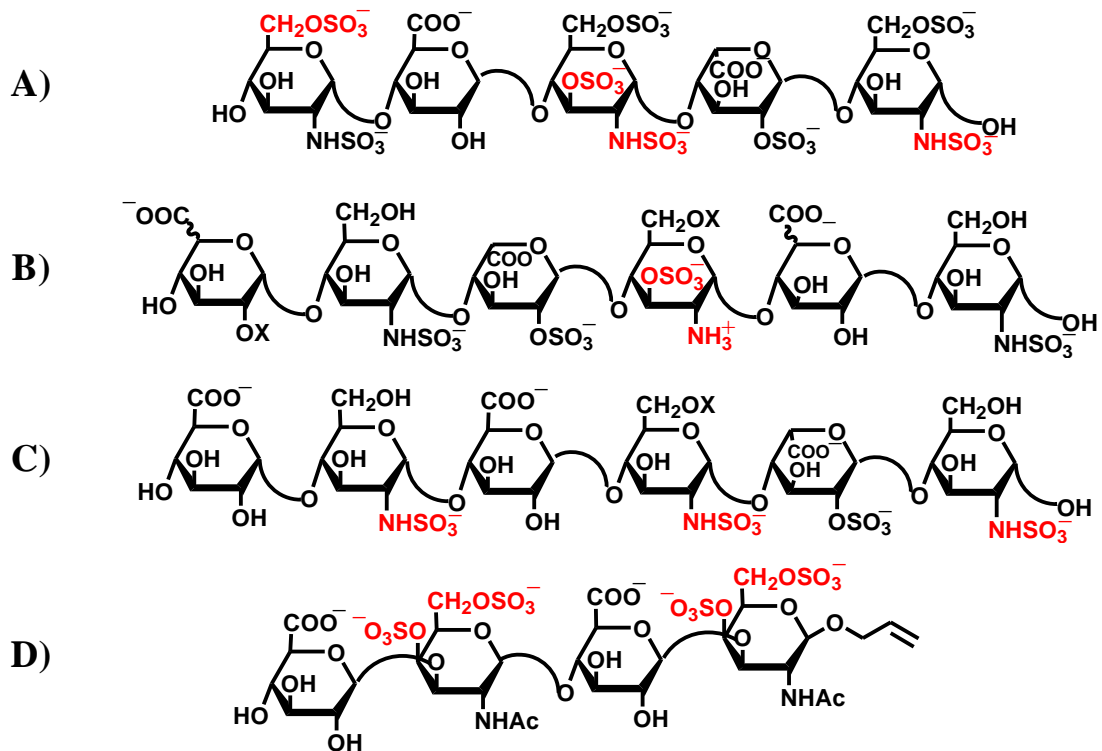
Figure 6. Specific Protein-Binding GAG Sequences. Critical sulfate group are highlighted in red. X = H or OSO_3^-

As little as a hexasaccharide is required to define the binding domain for FGF-2 (Figure 6C).[149-151] In addition to N-sulfation of the GlcNp units, the one IdoAp disaccharide is required for binding. It is known that HS serves to bridge FGF-2 and its receptor to form an active signaling complex in a 2:2:1 (FGF-2:FGF-2R:GAG) stoichiometry (see section 1.1.2). Interestingly, while 6-O-sulfation is not required for binding to FGF-2, it is an essential requirement for HS binding to the FGF-2 receptor. Thus, 6-O-desulfated sequences serve as antagonists by binding to FGF-2 and preventing receptor signaling.[152] More recently, a specific chondroitin sulfate (a type of GAG) tetrasaccharide that

recognizes the neuronal protein midkine has been reported (Figure 6D).[153] The authors have established through chemical synthesis of analogs that the distribution of sulfate groups on the tetrasaccharide is critical to activity.

Over-all, GAGs are versatile molecules that modulate biological functions through either non-specific or specific interactions with proteins. Given the widespread roles of GAGs in biology, specific GAG sequences provide an opportunity for the development of GAG-based molecules to treat a wide range of disorders.

1.5. Developing GAG-Based Therapeutics

Several challenges in drug development need to be overcome in order to exploit the vast potential of complex GAGs and clarify important structure-activity relationships. First, methods to identify specific GAG sequences from a complex mixture of sequences need to be developed. Given the structural complexity of GAGs and difficulties of isolation of small amount of HS from cell surfaces, this represents a monumental task. Previous methods have circumvented the problem of HS isolation by relying on heparin, which may be readily extracted from porcine intestinal mucosa, as a substitute. Thus, affinity column chromatography of heparin oligosaccharides is the most common method for identifying specific GAG sequences. Although this method has successfully resulted in the identification of the antithrombin-binding H5 sequence, it is clearly limited to the sequences present in heparin, which largely consists of a single, repeating disaccharide. In order to exploit the structural diversity in GAGs, affinity chromatography of HS rather than heparin needs to be carried out. Yet, this method is not practical since it requires

significant quantities of isolated HS. Furthermore, unlike DNA, small amounts of isolated HS cannot be amplified because of the complex GAG biosynthetic machinery. Recently, Sasisekharan and co-workers reported a mass spectrometric-based method to sequence small amounts (pmol) of isolated HS.[154,155] This method involved the immobilization of the target protein on a hydrophobic surface followed by the application of a HS mixture. Non-specific sequences were removed by salt washes and any adherent specific sequences were structurally characterized using MALDI-MS. This work is novel and has the potential to expedite the identification of specific GAG sequences. It is important to point out that the main drawback of a HS isolation/sequencing approach is the impending risk that the isolated HS sample may not contain the sequences responsible for *in vivo* activity because of the variability of HS structure with tissue type of origin.

Synthesis of a library of HS sequences followed by high-throughput screening represents a second approach to develop GAG-based therapeutics. Chemical synthesis of GAGs remains a formidable frontier and poses several challenges.[156] While initial efforts aimed at targeting individual sequences, current efforts are modular and are geared towards the synthesis of different sequences from a common set of building blocks. The two primary obstacles in the development of an efficient modular synthetic approach are: 1) restricted access to the non-natural *L*-iduronic acid monosaccharides, which are often prone to epimerization along the synthetic pathway, and 2) stereoselective formation of the inter-glycosidic linkages.

Automated, solid-supported synthesis of GAGs represents an attractive avenue to gain rapid access to HS libraries. While automated carbohydrate synthesis is a nascent

field, it is a rapidly burgeoning one too.[157] Seeberger and co-workers have achieved the automated synthesis of a dodecasaccharide using a modified Applied Biosystems 443 peptide synthesizer as the first prototype instrument.[157] These workers have also assembled a hexasaccharide-based malarial antigen using a semi-automated approach. To date, no GAG sequences have been assembled using automated, solid-supported synthesis. Another innovative approach towards the synthesis of GAG libraries is the programmable one-pot strategy that Chi-Huey Wong and co-workers have developed.[158] This approach relies on the sequential one-pot coupling of different monosaccharides that differ in their anomeric reactivity to yield a single sequence. Using this approach, a HS pentasaccharide was rapidly assembled.

Chemo-enzymatic approaches to gain access to HS libraries have also been reported.[159-161] These approaches involve chemical desulfation and/or deacetylation of heparin or heparosan polysaccharides followed by resulfation using O-sulfotransferase enzymes. By using different combinations of O-sulfotransferase enymes, libraries of polysaccharides may be generated. This approach can also be modified to generate oligosaccharides by cleaving heparin chains using heparanase before the resulfation step. While it is possible to generate milligram quantities of polysaccharides, chemo-enzymatic approaches provide limited information on structure-activity relationships because the enzymatic reactions do not proceed to completion and perform similar modifications across the entire polysaccharide chain. Nevertheless, chemo-enzymatic methods provide a powerful approach to further the development of GAG-based therapeutics and may serve to complement total synthesis efforts.

So far as high throughput screening of GAG libraries is concerned, several groups have reported carbohydrate microarrays to analyze carbohydrate-protein interactions and a few of these reports pertain to GAGs.[162-166] The miniature array format permits detection of multiple binding events simultaneously and requires minimal amount of carbohydrate and protein. While these methods provide an opportunity for the rapid interrogation of GAG-protein interactions, they are limited by the availability of GAG libraries.

In conclusion, several different experimental approaches to develop GAG-based therapeutics are being pursued. It is worth mentioning that no theoretical approaches to identify specific GAG sequences have been developed. Significant hurdles need to be surmounted at the present time before an era of GAG-based therapeutics ensues. This is evident from the fact that only one GAG sequence (the H5 sequence) is currently in the clinic.

DEVELOPING NON-SACCHARIDE MIMICS OF SPECIFIC GAG SEQUENCES

Our research group questioned the fundamental assumption that the saccharide skeleton is essential for the activity of specific GAG sequences. We hypothesized that specific GAG sequences contain structural redundancies that complicate drug development. In addition, non-saccharide heparin/HS mimics may provide several advantages over the GAG skeleton: 1) ease of laboratory-scale chemical synthesis and amenability to industrial-scale production to meet growing market demands; 2) possibility of oral delivery due to enhanced hydrophobic character when compared to GAGs; 3) possibility of additional non-ionic binding energy and enhanced specificity to the target protein; and 4) ability to modulate responses in either an agonist or an antagonist manner. In the pursuit of a simplified yet effective approach to capitalize on the diversity of GAGs, non-saccharide mimics of the H5 sequence were designed and evaluated.[167]

As described earlier, the H5 sequence binds to the serpin antithrombin (AT) in the pentasaccharide-binding site and induces a major conformational change that dramatically enhances its rate of factor-Xa inhibition (section 1.1.1). Desai and co-workers have shown through structure-activity relationship studies that while residues D, E, F, G, and H of the

24
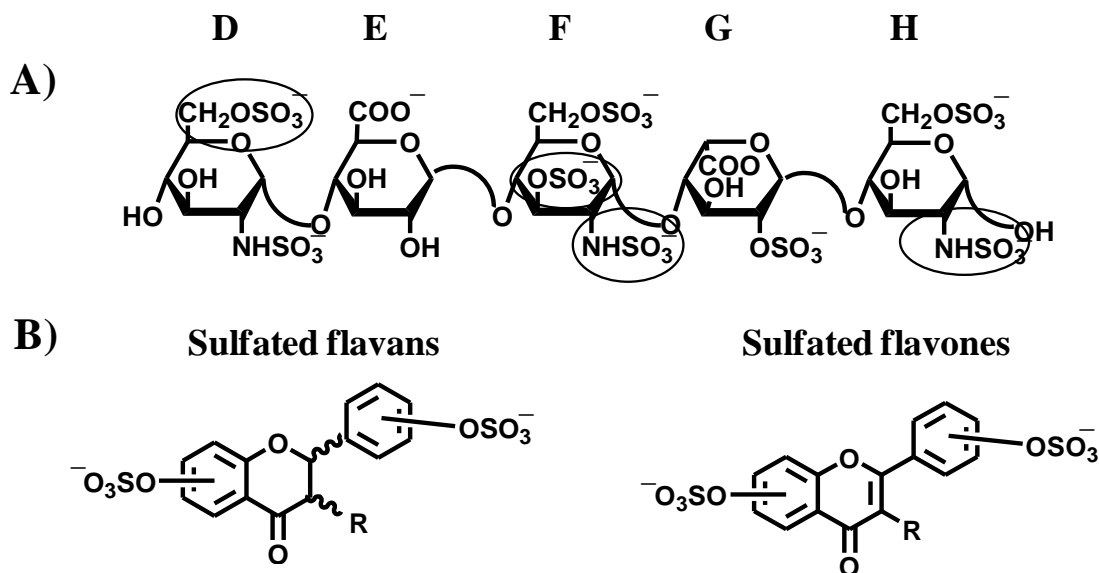
Figure 7. Structures of the H5 sequences (A), and non-saccharide DEF mimics (B). Critical sulfate groups of the H5 sequence are encircled. While the H_{2S} group contributes to affinity only, the D_{2S}, F_{2S} and F_{3S} groups are required for both high affinity and activation.

H5 sequences are required for high-affinity binding (50 nM) and complete activation (300-fold) of antithrombin, residues D, E and F can bring about full activation at 1000- fold higher concentrations (Figure 7A).[168] Thus, the H_{2S} group is important for high affinity interaction but does not contribute to AT activation.[*] Using the trisaccharide DEF as a template, Desai et al. designed small, non-saccharide molecules called sulfated flavans using hydropathic interaction (HINT)[169] analysis (Figure 7B). These molecules were 18 and 80 µM and weakly activate the inhibitor (~10-fold). Competitive binding experiments

---

[*] The importance of the H_{2S} group has been determined by the synthesis of an analog that is devoid of this group (van Boeckel, C. A. A; Petitou, M. *Angew. Chem. Intl. Ed. Engl.*, **1993**, *32*, 1671-1818)

readily synthesized in one step from commercially available flavonoids and biochemically evaluated.[170] The results indicated that sulfated flavans bind to AT with an affinity betweensuggested that the non-saccharide designs by-pass the targeted pentasaccharide-binding site in AT and bind in the adjoining extended heparin-binding site.[171] This result explains the weak activation of AT. In order to determine structure-activity relationships several analogs of sulfated flavans including sulfated flavones were synthesized (Figure 7B). However, initial synthetic efforts towards sulfated flavones were unproductive because these non-saccharide designs possess significantly greater charge density than GAGs.[172] Hence an alternate synthetic route to sulfated flavones that involved reductive cleavage of 2,2,2-trichloroethyloxysulfonyl-protected flavonoids was developed (Figure 8).[173] Using this method, six sulfated flavones bearing different sulfate group distributions were synthesized and evaluated. The results indicated that the activity of these molecules were comparable to sulfated flavans and not critically dependent on the location of sulfate groups.[173] Thus, sulfated flavans and flavones bind non-specifically to the smaller extended-heparin binding site in AT and result in weak activation.

To improve upon the activation potential of non-saccharide mimics, it was hypothesized that increasing the length of these molecules by the addition of a linker between the bicyclic and unicyclic flavonoid rings may favor binding to the larger pentasaccharide-binding site. Taking synthetic feasibility into consideration, HINT analysis was used to arrive at the tetrahydroisoquinoline-based design IAS$_5$ (Figure 9). When compared to sulfated flavonoids, this molecule contains a modified bicylcic ring, a

TCE = 2,2,2-trichloroethyloxysulfonyl (CCl$_3$CH$_2$OSO$_2$-)
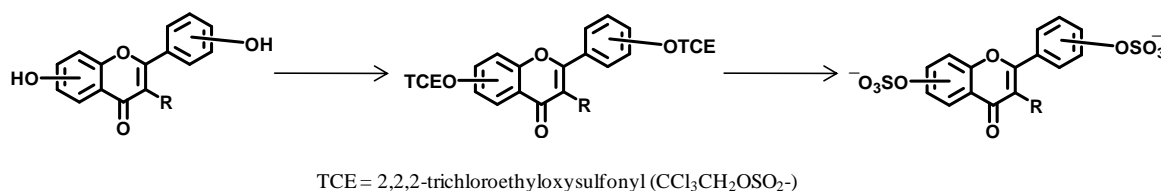
Figure 8. TCE-protection-deprotection strategy for the synthesis of sulfated flavones

one-carbon linker between bicyclic and unicyclic rings and an extra negative charge in the form of a carobxylate group. Molecular modeling was used to confirm that there was a close correspondence between the sulfate and carboxylate groups of a low energy

Figure 9. Structure of IAS$_5$
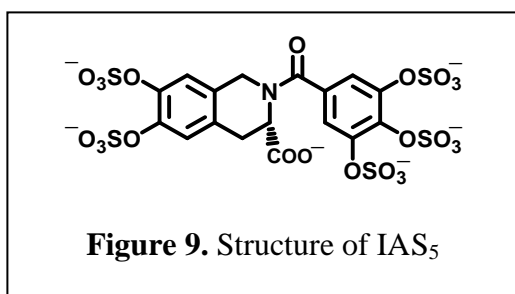
conformation of IAS$_5$ and that of the trisaccharide DEF. Despite exhaustive attempts to synthesize IAS$_5$ from commercially available precursors, IAS$_5$ remained intractable to chemical synthesis because of problems with sulfation and introduction of a free carboxylate group. In this regard, both traditional sulfation and TCE protection-deprotection failed.

2.1. Specific Aims of the Research Project

Taking into account the problems encountered with the chemical synthesis of small sulfated non-saccharide molecules and the lack of computational approaches to identify specific GAG sequences (section 1.5), a set of specific aims were devised as follows –

Aim 1) Develop a rapid and efficient synthetic method for highly sulfated small molecules

Aim 2) Synthesize and biochemically characterize the potential non-saccharide AT activator, IAS5, and its analogs

Aim 3) Develop a computational method to identify specific GAG sequences that may bind to any given GAG-binding protein

Aim 4) Use the method developed in aim 3 to design non-saccharide molecules that may target the pentasaccharide-binding site in antithrombin

Aim 5) Synthesize and biochemically characterize the non-saccharide designs from aim 4

MICROWAVE-BASED SYNTHESIS OF HIGHLY SULFATED
SMALL MOLECULES

3.1. Introduction

Recent work in our laboratory shows that designed highly sulfated, aromatic, small organic molecules possess interesting physico-chemical and biological properties.[167,170,171,173] Biochemically, these molecules form multiple ionic as well as non-ionic interactions, which form the backbone of most protein-recognition elements. Structurally, these represent mimics of glycosaminoglycans (GAGs), which are increasingly being recognized as modulators of key physiological functions,[2,3] while toxicologically, the sulfated structure represents a highly water-soluble, already-metabolized form that is expected to possess minimal toxicity. Despite these novel features, highly sulfated organic molecules remain largely unexplored.

A major limitation in exploring these novel structures is their challenging synthesis. Nearly all small organic sulfates reported in the literature are mono- or di-sulfated molecules,[174-178] typically prepared using sulfur trioxide complexes with amines in a highly polar solvent (DMF or DMA). Sulfation of such organic scaffolds may require as many as 13 hrs and temperatures as high as 95 °C in the presence of a large excess of the sulfating

29

complexes,[174-178] while sugars, which contain multiple –OH groups, require reaction times in the range of 12 hrs to several days.[153,179-182]

Theoretically, this method could be extended for synthesis of highly sulfated drug-like molecules, yet practically it is a synthetic nightmare because these molecules possess significantly higher negative charge density.[172] The major challenge is driving the reaction to completion in order to sulfate all available hydroxyl groups (alcoholic or phenolic) on the substrate. As the number of –OH groups increase on a small scaffold, sulfation becomes progressively more difficult because of anion crowding, resulting in numerous partially sulfated side-products.

A further challenge is the isolation of the chemically pure per-sulfated product, which requires aqueous isolation techniques. Yields in the range of 11 and 100% have been reported, yet the presence of inorganic salts arising from the use of buffers and salts leads to significant inconsistencies and inaccuracies. Additionally, instability of highly anionic products introduces limitations on reaction times and temperature.[153] This is likely to be especially true for highly sulfated, aromatic, small organic molecules, which are expected to be less stable than the saccharide scaffolds.

To avoid these problems with one-step sulfation, we recently synthesized some small aromatic per-sulfated structures using a two-step approach involving the 2,2,2-trichloroethyl protecting group.[173] The two-step protection-deprotection protocol resolved some of the problems of the direct sulfation approach, yet required careful real-time monitoring of the reaction by RP-HPLC to prevent product degradation and was not particularly applicable to substrates that were acid and/or metal sensitive. These limitations

led us to seek an alternative sulfation approach, which can be rapid, efficient, and widely
applicable to a number of poly-hydroxy scaffolds. We hypothesized that significant rate
enhancements are likely to be achieved using microwaves, especially because the ionic
sulfated product may couple to microwaves through ionic conduction, e.g., in CH$_3$CN.[183]
CH$_3$CN was chosen as the solvent over the commonly used DMF because a) it can be
evaporated at lower temperature (thus aiding isolation) and b) it was likely to solubilize the
per-sulfated product with an amine counter-ion. We also hypothesized that introducing free
base in the reaction mixture should promote the difficult per-sulfation reaction.

3.2. Results

We began our studies on microwave-aided sulfation using the polyhydroxyl
precursor **1** (see Table 2).* This structure was chosen for its resemblance to the targeted
molecule IAS5, which was designed to mimic the trisaccharide DEF of the antithrombin-
activating H5 sequence (see Figure 9).

Sulfation of **1** with SO$_3$●Me$_3$N complex (6 equivalents per –OH group) at 100 °C
in the absence of free base gave only 4.7 % of per-sulfated product **1s** in 20 minutes (Table
2, entry 1).** Inclusion of 1 equivalent of free Et$_3$N per –OH group resulted in 13.5 %
conversion (entry 2), while 79.8 % of **1s** was formed with 5 equivalents of Et$_3$N (entry 3).

---

* Compound **1** was synthesized in two steps from commercially available materials. See Chapter 4 for details.

** RP-HPLC profile showed peaks from 4.3-6.0 min, in addition to one at 9.0 min. The peak at 4.3 was subsequently isolated after optimization of conditions and determined to be per-sulfated (**1s**). The peak at 9.0 min was identified as **1** by comparison with synthetically pure sample. Conversions (%) were determined by area normalization.

Table 2. Optimization of microwave-assisted sulfation of tetrahydroisoquinoline derivative **1**
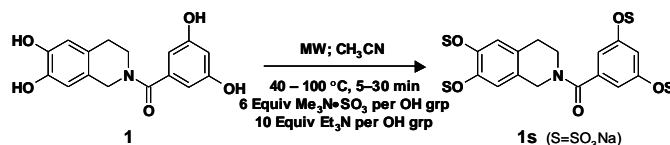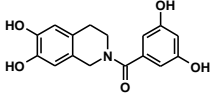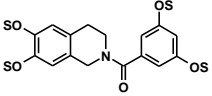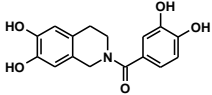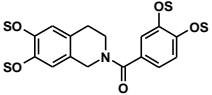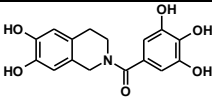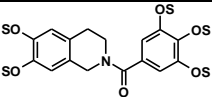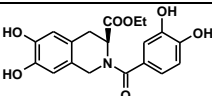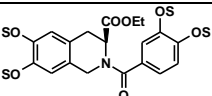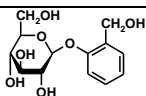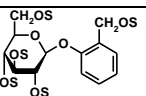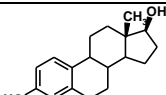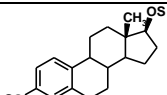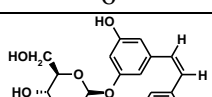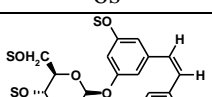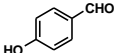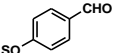
	Time (min)	Temp (°C)	Additional Conditions	HPLC Yield (%)
1	20	100	No base	4.7
2	20	100	1 equiv. Et$_3$N per OH grp.	13.5
3	20	100	5 equiv. Et$_3$N per OH grp.	79.8
4	20	100	—[a]	80.0
5	10	100	1 equiv. SO$_3$•Me$_3$N per OH grp.	14.5
6	10	100	3 equiv. SO$_3$•Me$_3$N per OH grp.	46.1
7	10	100	—[a]	47.4
8	10	100	9 equiv. SO$_3$•Me$_3$N per OH grp.	79.5
9	10	100	12 equiv. SO$_3$•Me$_3$N per OH grp.	80.7
10	10	100	DMF as solvent	17.2
11	10	100	CH$_3$NO$_2$ as solvent	23.2
12	10	100	With 6 equiv SO$_3$•py/OH grp and 10 equiv py/OH grp as base.	82.3
13	10	40	—[a]	0
14	10	70	—[a]	18.2
15	10	120	—[a]	90.8
16	30	100	—[a]	91.2

[a] Reaction conditions here are as listed above with no other modifications

Further increase to 10 equivalents Et_3N per –OH group had a negligible increase in the yield of **1s** (entry 4). Increasing the proportion of $SO_3 \bullet Me_3N$ per –OH group from 1 to 9 molar equivalents (Table 2, entries 5-8) gradually increased the yield of the per-sulfated product from 14.5 to 79.5 %, while further increase to 12 equivalent was found to be not particularly advantageous (entry 9). Thus, 6 and 10 molar equivalents of the sulfating complex and base, respectively, were chosen for further studies.

To assess the effect of solvent, we chose to evaluate nitromethane and DMF, both of which are solvents with high dielectric constant and known to be microwave-friendly. While only 23.2 and 17.2 % of per-sulfated product **1s** was formed from **1** in 10 minutes at 100 °C in CH_3NO_2 and DMF, respectively, 47.4% of the product was formed in CH_3CN (Table 2, entries 10 and 11). Thus, our initial choice of CH_3CN proved to be optimal. To assess the effect of temperature and reaction time, sulfation was performed for 10–30 minutes at 40 to 120 °C. While 30 minutes were required to yield 91.2 % of **1s** at 100 °C, only 10 minutes were needed for 90.8 % conversion at 120 °C. In striking contrast, no product was detected at 40 °C within 10 minutes. Finally, $SO_3 \bullet py/py$ complex was found to give nearly twice as much per-sulfated product as $SO_3 \bullet Me_3N/Et_3N$ complex (entries 7 and 12) in 10 minutes at 100 °C. Since pyridine is ~ 10,000-fold weaker base in comparison to Et_3N, this result suggests general base catalysis as the predominant mechanism of sulfation rather than a process involving deprotonation of the substrate followed by nucleophilic attack.

Table 3. Microwave-assisted sulfation of poly-hydroxyl substrates

$$\text{Substrate} \xrightarrow[\substack{\text{Et}_3\text{N (10 equiv/OH), SO}_3\bullet\text{Me}_3\text{N} \\ \text{(6 equiv/OH), CH}_3\text{CN}}]{\text{Microwaves, 100 °C, 20-30 min}} \text{Product}$$
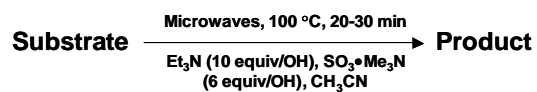
Entry	Substrate	Product [a]	Isolated Yield (%)
1	*(structure)* 1	*(structure)* 1s	85
2	*(structure)* 2	*(structure)* 2s	87
3	*(structure)* 3	*(structure)* 3s	54[b]
4	*(structure)* 4	*(structure)* 4s	74
5	*(structure)* 5	*(structure)* 5s	84[c]
6	*(structure)* 6	*(structure)* 6s	94[c]
7	*(structure)* 7	*(structure)* 7s	72[d]
8	*(structure)*	*(structure)*	97[c,e]

[a]S: SO_3Na. [b]With 9 equiv of $SO_3\bullet Me_3N$/OH group. [c]With $SO_3\bullet py$ (6–9 equiv/OH group) and pyridine as base. [d]Reaction conditions: 120°, 10 min, $SO_3\bullet py$ (12 equiv/OH group) and pyridine as base. [e]S = $SO_3\bullet pyH^+$

Appropriate control reactions in the absence of microwaves using two different substrates – **1** and **3** (entries 2 and 3 in Table 3) – at 60 °C in DMF with no free base showed poor product yields. For example, it took 24 hrs in the absence of microwaves to yield **1s** in 60% yield, while **3s** was not detected even after 24 hrs (entry 3, Table 3). These results highlight the importance of microwaves in achieving rapid per-sulfation.

Having optimized the reaction conditions, we assessed whether the method works for a variety of different substrates. Per-sulfation of **2** proceeded smoothly in a manner identical to **1** (Table 3).[24] More importantly, per-sulfation of **3**, containing the crowded 3,4,5-trihydroxy moiety, was achieved under microwave conditions in an isolated yield of 54 %, while the conventional procedure completely failed to give **3s**. Finally, microwave-assisted per-sulfation also works extremely well for substrates **5** through **8** containing one to six −OH groups. Interestingly, **5** and **6** gave a mixture of products with $SO_3 \bullet Me_3N$, but yielded the per-sulfated products with $SO_3 \bullet py$.

3.3. Discussion

Several points make the microwave-assisted synthetic protocol particularly attractive. A) The method appears to tolerate a range of functional groups including amide (Table 3, entries 1 through 4), ester (entry 4), aldehyde (entry 8) and double bond (entry 7). The relatively isolated yields (~70–95%) in each case make the reaction especially suitable for library construction. B) The method works equally well for substrates containing one −OH group and those that contain six −OH groups. This is important because the small size of these molecules introduces considerable anion−anion repulsion as the number of sulfate

groups increase. C) The method applies equally well to alcoholic and phenolic –OH groups, especially with $SO_3 \bullet py$ complex. D) The method provides high purity per-sulfated product that is readily isolated using an aqueous G10 filtration column. Typically, the purity of these highly water soluble, per-sulfated, small, organic molecules was found to be more than 95% using reverse polarity capillary electrophoresis. E) The method is particularly suitable for quantitative isolation of small amounts (<10 mg) of the per-sulfated products, but could be linearly scaled up at least 20-fold without affecting the yields to a significant extent.

It appears that the rate-accelerations achieved in our experiments is related to the phenomenon of microwave-induced dielectric heating[184] and the ability of acetonitrile to efficiently absorb microwave energy and convert it into heat. When the reaction mixture is irradiated with microwaves at a frequency of 2.45 GHz, the acetonitrile dipoles align in the applied electric field. As the applied field oscillates, the dipoles tend to realign with oscillating electric field and, in the process, lose energy in the form of heat (dielectric loss). If the solvent dipoles realign too quickly or do not have enough time to realign, no heating will occur. It appears that at 2.45 GHz, the frequency of most commercially available microwave system, the molecular dipole realignment time is optimal. Microwave irradiation of the sulfation reaction mixture therefore produces internal heating resulting in highly efficient heat transfer and consequently rapid reaction rates.

In summary, we have developed a rapid and high yielding microwave-based synthesis of variably functionalized, per-sulfated organic molecules. The protocol is

expected to greatly facilitate the construction of a library of per-sulfated, small organic molecules for screening as glycosaminoglycan mimetics.

3.4. Experimental Section

3.4.1. General Methods

All reactions sensitive to air or moisture were carried out under nitrogen atmosphere in oven-dried glassware. All reagent solutions unless otherwise noted were handled under an inert nitrogen atmosphere using syringe techniques. Anhydrous dichloromethane and acetonitrile were purchased from Sigma-Aldrich and Acros Organics, respectively, and were used without further drying. Trimethylamine-sulfur trioxide and pyridine-sulfur trioxide complexes were purchased from Alfa-Aesar and Fluka, respectively. All other reagents/chemicals were purchased from Sigma-Aldrich and were used as supplied. Analytical thin-layer chromatography (TLC) was performed using UNIPLATE^{TM} silica gel GHLF 250 μm pre-coated plates (ANALTECH, Newark, DE) that were analyzed by fluorescence (254 nm). Column chromatography was performed using silica gel (40-60 μm, 60 Å, Silicycle, Quebec, Canada) and the indicated technical grade solvents. After chromatography, solvents were evaporated using a Büchi rotary evaporator, followed by further treatment under high vacuum.

Microwave-based sulfation reactions were performed using a CEM-Discover (Matthews, NC) synthesizer in sealed reaction vessels (7 mL). The stirring parameter in the microwave synthesizer was set to "Hi-speed". The reaction mixture was ramped to 100 or

120 ℃ using the following power-temperature steps: 1) 50 W, r.t. – 80 ℃, and 2) 10 W, 80-100/120 ℃. The reaction vessel was simultaneously cooled using nitrogen (45 psi) to maintain the set temperature.

Sephadex G10 chromatography (de-salting) and SP Sephadex-Na chromatography (cation exchange) were performed using Flex columns (KIMBLE/KONTES, Vineland, NJ) of dimensions 170×1.5 cm and 75×1.5 cm, respectively. Cation exchange was performed with 30-fold excess of sodium ion equivalents. Samples were chromatographed at a controlled flow rate of 0.5 mL/min with water as eluent. Five mL fractions were collected and analyzed by RP-HPLC or capillary electrophoresis (see below).

^{1}H-NMR and ^{13}C-NMR were recorded on Varian Mercury-300 MHz or Varian Inova-400 MHz spectrometers in $CDCl_3$, DMSO-d6, CD_3OD or CD_3COCD_3. All signals are reported in ppm with the internal chloroform, DMSO, CD_3OD and CD_3COCD_3 signals at 7.26, 2.50, 3.31 and 2.05 ppm, respectively, as standards. The data is being reported as: chemical shifts (ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quadruplet, m = multiplet or unresolved, br = broad signal), coupling constant(s) (Hz), and integration

ESI mass spectra were recorded using a Micromass ZMD 4000 single quadrupole spectrometer. Samples were dissolved in methanol or methanol-acetonitrile (1:1) and infused at a rate 10 μL/min. Mass scans were obtained in the range 100-1300 amu at a scan rate of 400 amu/s. Ionization conditions were optimized for each compound to maximize ionization of the parent ion. The capillary voltage was varied between 3.0 and 4.5 V, while the cone voltage usually ranged from 20 to 80 V. For all experiments, the extractor voltage

was set to 4.0, the Rf lens voltage to 0.1 V, the source block temperature to 100 ºC and the desolvation temperature to 120 ºC.

Capillary electrophoresis using a Beckmann PACE/ MDQ unit was performed to test completion of sulfation reaction and assess isolated product purity. An uncoated fused silica capillary of 50 μm internal diameter and 32.5 cm effective length to the detector window was used. Samples were typically injected under a pressure of 0.5-1 psi for 5 s and detected spectrophotometrically using a 254 nm filter. Electrophoresis was performed under reverse polarity conditions at 25 °C and a constant voltage of 10 kV using 20 mM sodium phosphate, pH 2.7 or 4.3.

HPLC analysis was carried out on a Shimadzu chromatography system using Waters Atlantis dC18 column (5μ, 4.6×250 mm). The mobile phase consisted of a 100 mM sodium chloride-acetonitrile mixture(7:3 v/v) run at a constant flow rate of 0.5 mL/min. Analysis was carried out using a uv-vis detector at 254 nm.
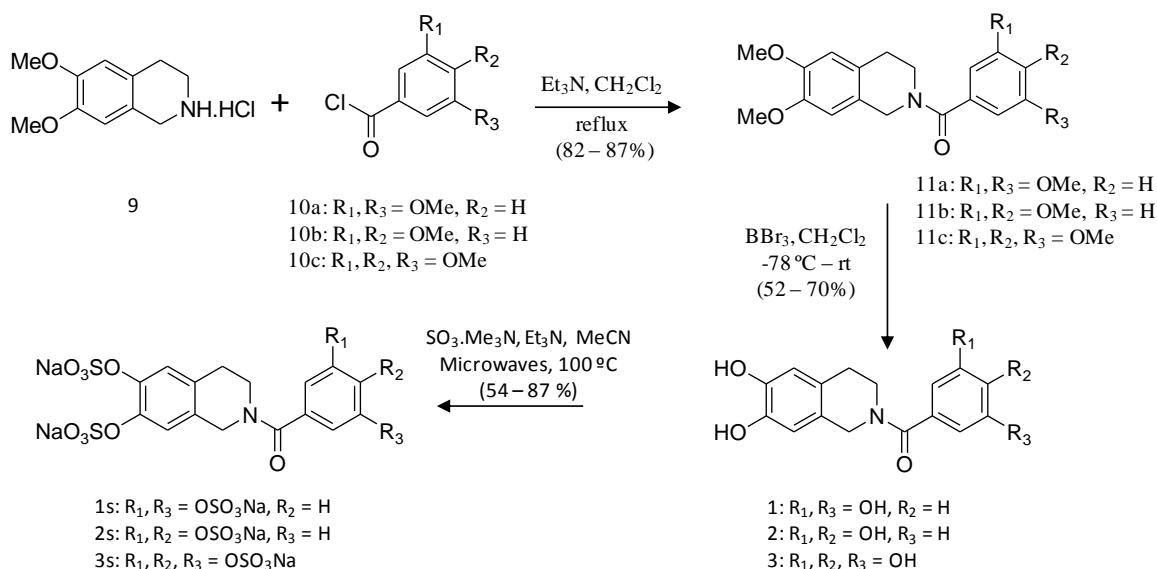
Figure 10. Synthesis of poly-sulfated tetrahydrosioquinoline derivatives **1s**-**3s**

3.4.2. Experimental procedures and spectral data

Synthesis of amides **11a** -**c**: To a stirred suspension of amine **9** (2g, 8.7 mmol) and triethylamine (6.1 mL, 43.5 mmol) in dichloromethane (40 mL) at 0 ºC, was added acid chloride **10a-c** (9.14 mmol, 1.05 equiv.). The reaction mixture was allowed to warm to room temperature and refluxed. After 4 hrs, the reaction mixture was diluted with dichloromethane (50 mL), washed with 0.5 N HCl (3×50 mL) and potassium carbonate (3×50 mL), dried over anhydrous Na_2SO_4, and concentrated *in vacuo* to obtain a colorless oil (82 **-** 87 %). **11a:** ^{1}H-NMR (300 MHz, CD_3COCD_3): δ 6.59 – 6.90 (m, 5H), 4.71 (br s, 2H), 3.64 – 3.84 (m, 14 H), 2.87 (br, 2H); ESI (+ve) *m/z* calcd for $C_{20}H_{23}NO_5$ [(M+H)[+]] 358.17, found 358.3; **11b:** ^{1}H NMR (300 MHz, CD_3COCD_3) δ 6.87 – 6.92 (m, 3H), 6.60

(s, 2H), 4.5 (s, 2H, CH_2), 3.63 – 3.72 (m, 14 H), 2.67 (t, J = 5.7 Hz, 2H); ESI (+ve) m/z calcd for $C_{20}H_{23}NO_5$ [(M+H)$^+$] 358.17, found 358.1; **11c:** ^{1}H-NMR (300 MHz, CD_3COCD_3): δ 6.77 – 6.79 (m, 4H), 4.68 (s, 2H), 3.77 – 3.88 (m, 17 H), 2. 84 (t, J = 5.7 Hz, 2H); ESI (+ve) m/z calcd for $C_{21}H_{25}NO_6$ [(M+H)$^+$] 388.18, found 388.1

Polyphenols **1-3**: To a stirred solution of the amide (7.0 – 7.5 mmol) in dichloromethane (80 mL) at -78 ºC, was added BBr$_3$ (36 – 42 mL of 1M solution in CH_2Cl_2, 1.2 equiv per OMe group) under N$_2$ over 15 minutes. After stirring for 12 hrs at rt, the reaction was quenched at 0 °C with MeOH (10 mL) and water (10 mL). The reaction mixture was partitioned between EtOAc (220 mL) and 2N HCl (50 mL). The aqueous layer was diluted with brine (50 mL) and washed with EtOAc (6 x 50 mL). The combined organic layer was dried over Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was purified by silica gel chromatography (Hexanes/EtOAc = 1:1, 1:4, 1:4.5, 0:1) to give a yellow solid (52 - 70%). **1:** ^{1}H NMR (300 MHz, CD$_3$OD) δ: 6.30 – 6.61 (m, 5H, isomers I & II), 4.66 (s, 2H, isomer I), 4.44 (s, 2H, isomer II), 3.88 (t, J = 6.0 Hz, 2H, isomer II), 3.62 (t, J = 6.0 Hz, 2H, isomer I), 2.80 (t, J = 6.0 Hz, 2H, isomer II), 2.73 (t, J = 6.0 Hz, 2H, isomer I); ESI (+ve) m/z calcd for $C_{16}H_{15}NO_5$ [(M+H)$^+$] 302.10, found 302.2, ESI (-ve) calcd for $C_{16}H_{15}NO_5$ [(M-H)$^-$] 300.09, found 300.07; **2:** 1H NMR (300 MHz, CD$_3$OD): δ 6.40 - 6.91 (m, 5H), 4.63 (br s, 2H), 3.68 – 3.86 (m, 2H), 2.77 (s, 2H); ESI (+ve) calcd m/z for $C_{16}H_{15}NO_5$ [(M+H)$^+$] 302.10, found 302.16, ESI (-ve) calcd for $C_{16}H_{15}NO_5$ [(M-H)$^-$] 300.09, found 300.1; **3:** ^{1}H NMR (400 MHz, CD$_3$OD): δ 6.54 (s, 2H), 6.44 (s, 2H), 4.48 – 4.56 (m, 2H), 3.62 – 3.78 (m, 2H), 2.72 (s, 2H); ESI (+ve) m/z calcd for $C_{16}H_{15}NO_6$

[(M+H)$^+$] 318.10, found 318.0, ESI (-ve) calcd for C$_{16}$H$_{15}$NO$_6$ [(M-H)$^-$] 316.08, found 316.0

Per-sulfates **1s and 2s**: To a stirred solution of the poly-alcohol (20 mg, 0.066 mmol) in MeCN (1 mL) at rt, Et$_3$N (0.4 mL, 2.9 mmol) and Me$_3$N.SO$_3$ (220 mg, 1.6 mmol) was added. The reaction vessel was sealed and micro-waved for 20 minutes at 100 °C. The reaction was repeated for 4 times and the reaction mixture was pooled together. The MeCN layer was decanted and pooled, while the residue from each tube was washed with MeCN (5 mL) and centrifuged. The combined MeCN layers were concentrated *in vacuo*. Water (5 mL) was added to the residue and stirred for 10 min. The water layer was concentrated to approximately 2 mL, loaded onto a Sephadex G10 column (~ 160 cm) and chromatographed using water as eluent. Fractions were combined based on RP-HPLC profiles, concentrated and re-loaded onto a SP Sephadex C25 column for sodium exchange. Appropriate fractions were pooled, concentrated *in vacuo* and lyophilized to obtain a white powder (84 - 87 %). **1s:** ^{1}H NMR (DMSO, 400 MHz) δ: 7.29 – 7.30 (m, 2H), 6.94 – 6.97 (m, 3H), 4.58 (s, 2H, isomer I), 4.48 (s, 2H, isomer II), 3.58 (s, 2H, isomer II), 3.50 (s, 2H, isomer I), 2.66 (br, 2H, isomer I & II); ESI (-ve) *m/z* calcd for C$_{16}$H$_{11}$NNa$_4$O$_{17}$S$_4$ [(M-Na)$^-$] 685.86, found 686.1; **2s:** ^{1}H NMR (DMSO, 400 MHz) δ: 7.65 (d, *J* = 2.4 Hz, 1H), 7.57 (d, *J* = 8.4 Hz, 1H), 7.29 (s, 2H), 6.99 (dd, *J* = 8.4, 1.6 Hz, 1H), 4.54 (s, 2H), 3.70 (br, 2H), 2.69 (t, *J* = 4.8 Hz, 2H); ESI (-ve) *m/z* calcd for C$_{16}$H$_{11}$NNa$_4$O$_{17}$S$_4$ [(M-Na)$^-$] 685.86, found 686.0

Per-sulfate **3s**: To a stirred solution of the poly-alcohol (20 mg, 0.063 mmol) in MeCN (1.3 mL) at rt, Et$_3$N (0.5 mL, 3.6 mmol) and Me$_3$N.SO$_3$ (390 mg, 2.8 mmol) was

added. The reaction vessel was sealed and microwaved for 30 minutes at 100 ℃. The reaction was repeated for 4 times for scale up. The product **3s** (121 mg, 54 %) was isolated according to the above procedure for **1s** and **2s**. **3s:** ^{1}H NMR (DMSO, 400 MHz) δ: 7.37 (s, 2H), 7.29 (s, 2H), 4.54 (s, 2H), 3.53 (s, 2H), 2.68 (s, 2H); ESI (-ve) *m/z* calcd for $C_{16}H_{10}NNa_5O_{21}S_5$ [(M-Na)[-]] 803.79, found 804.1
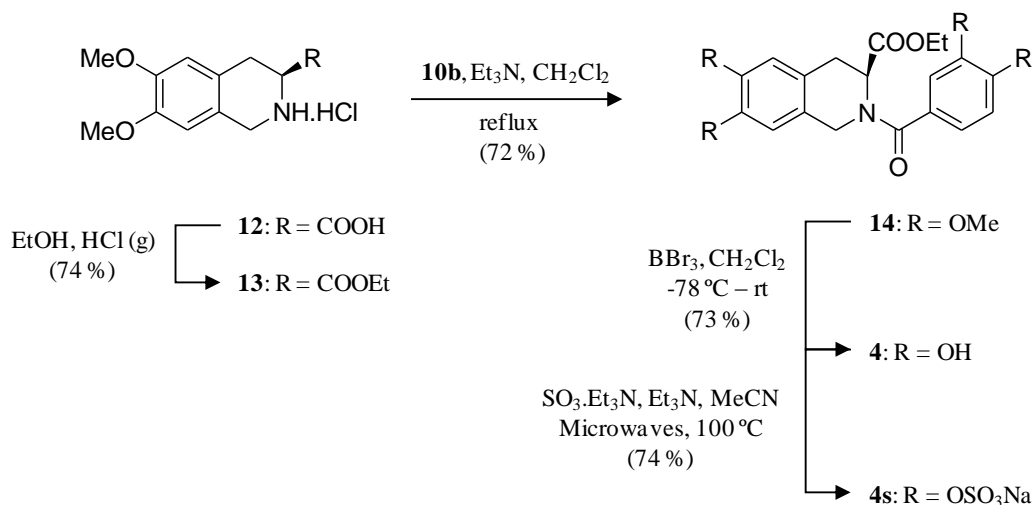
Figure 11. Synthesis of poly-sulfate **4s (IES4)**

Ester **13** — To a stirred solution of amino acid **12** (2 g, 7.32 mmol) in EtOH (50 mL) at rt, HCl (g) was passed for 2 minutes and the reaction was refluxed. After 24 hrs, the reaction was brought to rt and the solvent was evaporated off. Ethyl acetate (50 mL) was added to the residue and extracted with 5 % K_2CO_3 solution (3 x 25 mL) and water (2 x 20 mL). The organic layer was dried (Na_2SO_4) and evaporated to obtain a colorless oil (1.4 g, 72 %). ^{1}H NMR (CDCl₃, 400 MHz) δ: 6.59 (s, 1H), 6.51 (s, 1H), 4.22 (q, *J* = 7.2 Hz, 2H),

4.03 (s, 2H), 3.84 (s, 3H), 3.82 (s, 3H), 3.68 (dd, J = 10.4, 4.4 Hz, 1H), 2.98 (dd, J = 16, 4.4 Hz, 1H), 2.86 (dd, J = 15.6, 10 Hz, 1H), 1.29 (t, J = 7.2 Hz, 3H)

Amide **14:** The procedure for the synthesis of amides **11a-c** was used to prepare **14** (72%) ^{1}H NMR (DMSO, 400 MHz) δ: 6.69 – 7.03 (m, 5H), 4.81 – 5.14 (m, 1H, isomers I - III), 4.27 – 4.57 (m, 2H, isomers I – III), 3.98 – 4.04 (m, 2H, isomers I – III), 3.62 – 3.78 (m, 12H, isomers I – III), 3.09 – 3.32 (m, 2H, isomers I – III), 1.00 – 1.11 (m, 3H, isomers I – III); ESI (+ve) m/z calcd for $C_{23}H_{27}NO_7$ [(M+H)$^+$] 430.18, found 430.4

Poly-phenol **4**: The procedure for the synthesis of polyphenols **1-3** was to prepare **4** from amide **14** (73%). ^{1}H NMR (DMSO, 400 MHz) δ: 8.78 – 9.35 (m), 6.35 – 6.81 (m, 5H), 4.69 – 4.99 (m, 1H, isomers I - III), 4.17 – 4.43 (m, 2H, isomers I – III), 3.50 – 3.57 (m, 2H, isomers I – III), 2.89 – 3.02 (m, 2H, isomers I – III), 1.00 – 1.11 (m, 3H, isomers I – III)

Per-sulfate **4s (IES4)**: The procedure for the synthesis of poly-sulfates **1s** and **2s** was used to prepare **4s** from polyphenol **4** (74%). ^{1}H NMR (DMSO, 400 MHz): δ 6.96 – 7.70 (m, 5H), 4.83 – 5.16 (m, 1H isomers I - III), 4.18 – 4.44 (m, 2H, isomers I – III), 3.55 – 3.61 (m, 2H, isomers I – III), 2.99 – 3.13 (m, 2H, isomers I – III), 1.08 – 1.16 (m, 3H, isomers I – III); ESI (-ve) m/z calcd for $C_{19}H_{15}NNa_4O_{19}S_4$ [(M-Na)$^-$] 757.88, found 757.5

Salicin per-sulfate **5s** (Figure 12): To a stirred suspension of salicin (**5**, 20 mg, 0.07 mmol) in MeCN (1 mL) at rt, pyridine (0.3 mL, 3.8 mmol) was added to form a clear solution. $C_5H_5N.SO_3$ (500 mg, 3.14 mmol) was added and the reaction vessel was sealed and micro-waved for 30 minutes at 100 ℃. The reaction was repeated for 4 times. The reaction
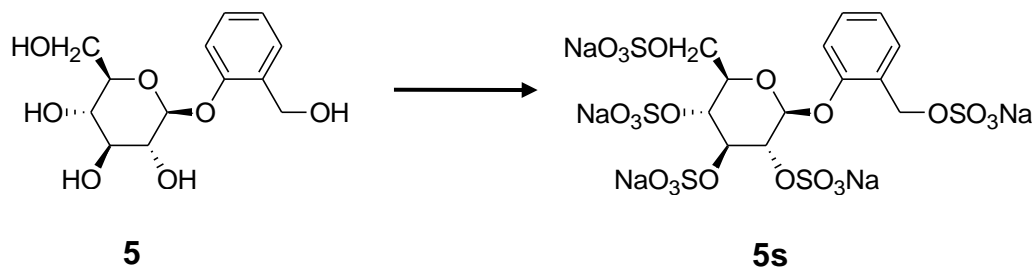
Figure 12. Synthesis of poly-sulfate **5s**

mixture was quenched with methanol (2 mL), pooled and concentrated *in vacuo*. The

residue was taken up in water (1 mL) and chromatographed following the protocol

established for **1s** to obtain **5s** (236.8 mg, 84 %). ^{1}H NMR (DMSO, 400 MHz) δ: 7.33 –

7.35 (m, 1H), 7.16 – 7.20 (m, 1H), 6.93 – 6.97 (m, 1H), 6.88 – 6.90 (m, 1H), 5.24 (d, *J* =

5.2 Hz, 1H) (s, 2H), 3.53 (s, 2H), 2.68 (s, 2H); ESI (-ve) *m/z* calcd for $C_{16}H_{10}NNa_5O_{21}S_5$

[(M-Na)$^-$] 772.81, found
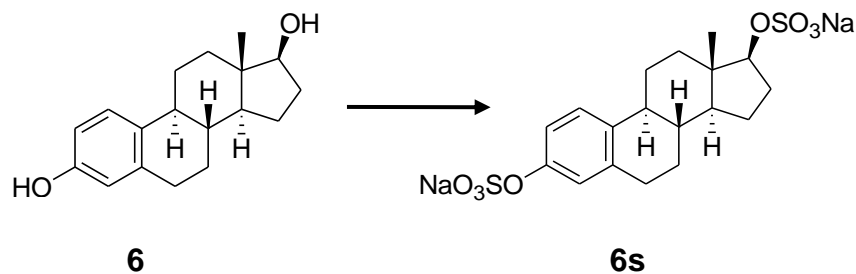
Figure 13. Synthesis of estradiol-3,17β-disulfate (**6s**)

772.7

Estradiol-3,17β-disulfate (**6s**): The procedure for the synthesis of salicin per-sulfate

5s was used to prepare **6s** from estradiol (94%). ^{1}H NMR (DMSO, 400 MHz) δ: 7.1 (d, *J* =

8.4 Hz, 1H), 6.81 – 6.83 (m, 2H), 4.02 (t, *J* = 8.0 Hz, 1H), 2.71 – 2.74 (m, 2H), 2.21 – 2.25

(m, 1H), 2.08 – 2.12 (m, 1H), 1.87 – 2.00 (m, 2H), 1.75 – 1.78 (m, 1H), 1.48 – 1.60 (m, 2H), 1.09 – 1.34 (6H), 0.66 (s, 3H); ESI (-ve) *m/z* calcd for $C_{18}H_{22}Na_2O_8S_2$ [(M-Na)⁻] 453.07, found 453.1
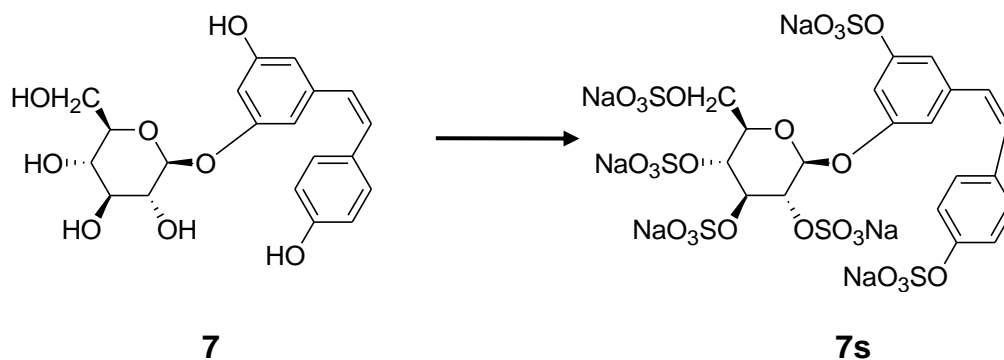
Figure 14. Synthesis of poly-sulfate **7s**

7s: The procedure for the synthesis of salicin per-sulfate **5s** was essentially used to prepare **7s** from stilbene **7**, except that 12 equiv. SO₃.py complex per –OH group was used and the reaction vessel was irradiated for 10 min at 120 °C (72%). ¹H NMR (DMSO, 400 MHz) δ: 7.56 (s, 1H), 7.54 (s, 1H), 6.99 – 7.10 (m, 7H), 6.61 (s, 1H), 5.36 (s, 1H), 4.63 (s, 1H), 4.59 (s, 1H), 4.39 (s, 1H), 4.14-4.18 (m, 1H), 4.08 (d, *J* = 8.8 Hz, 1H) 3.81 (t, *J* = 10 Hz, 1H); ESI (-ve) *m/z* calcd for $C_{20}H_{16}Na_6O_{26}S_6$ [(M-Na)⁻] 978.77, found 979.0

***p*-Hydroxy benzaldehyde *O*-sulfate** (**8s**) (Figure 15): The procedure for the synthesis of salicin per-sulfate **8** was essentially used to prepare **8s** from aldehyde **8**, except that the reaction vessel was irradiated for 10 min at 120 °C following which the product with pyridinium cation was directly isolated by lyophilization after G10 chromatography (97%,). ¹H NMR (DMSO, 400 MHz) δ: 9.76 (s, 1H), 8.94 – 8.96 (m, 2H), 8.62 – 8.68 (m,

1H), 8.09 – 8.14 (m, 2H), 7.72 – 7.75 (m, 2H), 6.92 – 6.95 (m, 2H); ESI (-ve) *m/z* calcd for

$C_{12}H_{11}NO_5S$ [(M-pyH$^+$)$^-$] 200.99, found 200.8
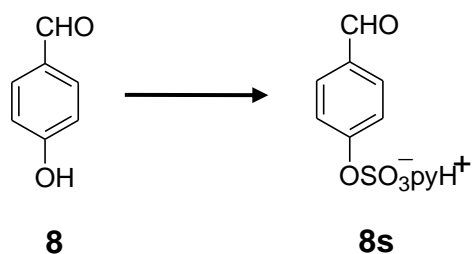
Figure 15. Synthesis if poly-sulfate **8s**

3.4.3. Capillary electropherograms of sulfated compounds 1s – 8s

The following electropherograms were recorded to assess the purity of synthesized poly-sulfated compounds **1s**-**8s** (see Table 3). Experimental conditions may be found in section 3.4.1).
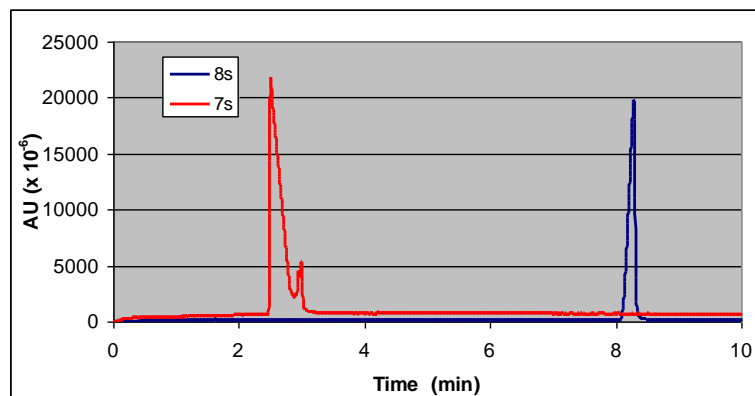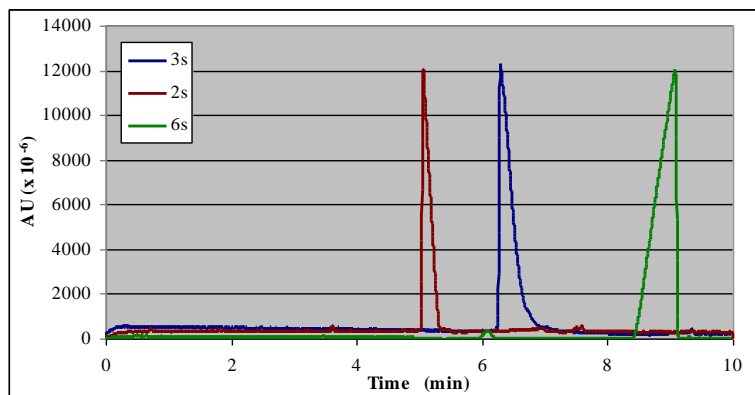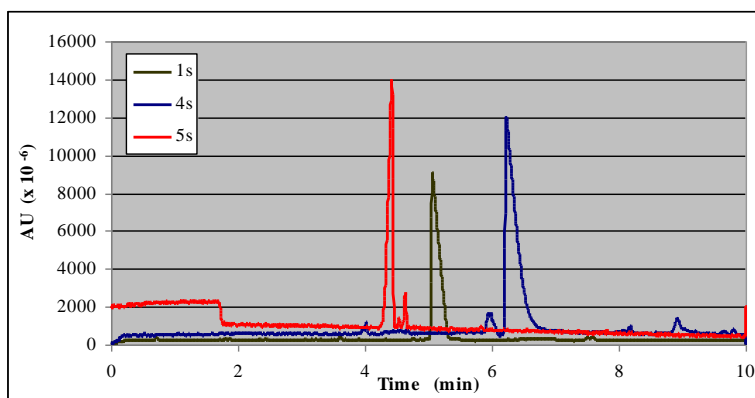
SYNTHESIS AND BIOLOGICAL EVALUATION OF DESIGNED ANTITHROMBIN ACTIVATOR IAS5 AND ANALOGS

4.1. Introduction

Antithrombin (AT)-based anticoagulants, first introduced in 1916,[6] include heparin, low molecular weight heparin (LMWH) and the recently (2001) introduced fondaparinux (FX),[185] and are all administered parenterally. Structurally, these agents belong to the glycosaminoglycan superfamily and are highly anionic molecules.[*] Their anticoagulant effect arises from binding to and activating the circulating coagulation inhibitor, AT .

The prototype of this group, heparin, suffers from several limitations: e.g., potential to cause bleeding, heparin-induced thrombocytopenia (HIT), osteoporosis, platelet function inhibition, and unpredictable dose-response effects.[7] While LMWH is gradually replacing heparin and is more therapeutically cost-effective, it still suffers from similar drawbacks.[186] The heterogeneity and polydispersity of both heparin and LMWH are the causes of the above mentioned side-effects. These side-effects are reduced by using the homogeneous heparin preparation — fondaparinux, which is a synthetic five-residue sequence based on

---

[*] Detailed information on the structure and conformation of heparin may be found in section 1.2

the naturally occurring H5 sequence (Figure 7). Compared to heparin or LMWH, result s from clinical trials with fondaparinux are more promising with a low probability of causing HIT but its long term efficacy and safety is yet to be ascertained.[187] In addition, fondaparinux does not interact with protamine sulfate, the heparin antidote, making it difficult to manage drug-induced bleeding. Nevertheless, the initial success of fondaparinux has validated factor Xa as a target for new AT-based anticoagulants.

The heparin-binding site in antithrombin (AT) is an engineering marvel. The binding site is located some 20 Å away from the reactive center loop (RCL) which is required for recognition and covalent inhibition of target proteases like thrombin and factor Xa. The heparin-binding site is formed by the positively charged residues of helices A and D of AT, and the polypeptide N-terminus.[118] Binding of the H5 sequence causes a conformational change that is relayed to the RCL resulting in enhanced exposure (Figure 16).[188] Specifically, H5 binding causes elongation of helix D (1-2 turns) which results in positive pressure on sheet A. This results in expulsion of the partially inserted RCL and enhanced recognition of target proteases. Importantly, the conformational change also results in the creation of an exosite in β-strand 3C that plays an important role in specific recognition of factor Xa.[189] Thus, conformational activation by H5 specifically enhances the rate of the AT-factor Xa reaction.

The H5-binding domain, called the pentasaccharide-binding site (PBS), is primarily formed by a positively charged triad comprised of Lys114, Lys125 and Arg129 (Figure 17). Biochemical site-directed mutagenesis studies suggest that these three residues
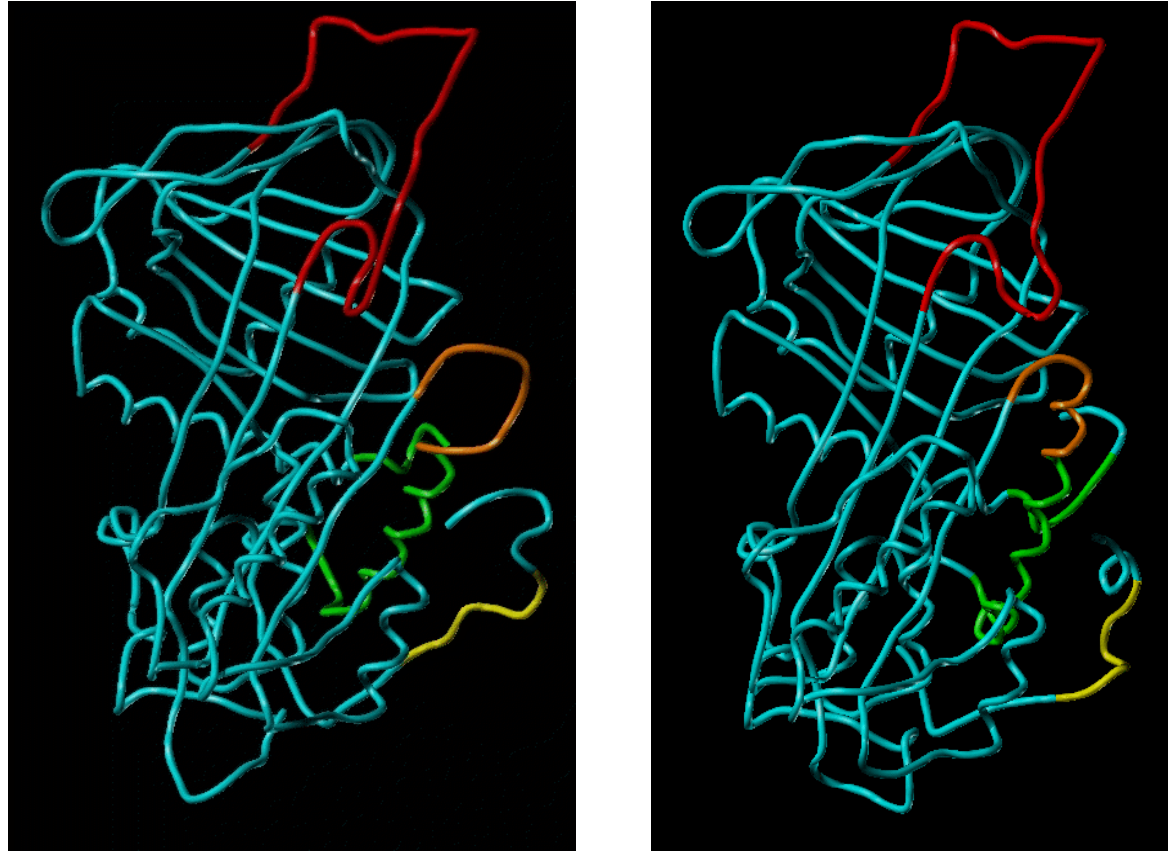
Figure 16. Structures of the native (left) and activated (right) conformations of antithrombin. Structural differences are highlighted as follows – red: reactive center loop (RCL), green: helix D of the heparin-binding site, orange: elongation of helix D on heparin binding, yellow: formation of a new helix P
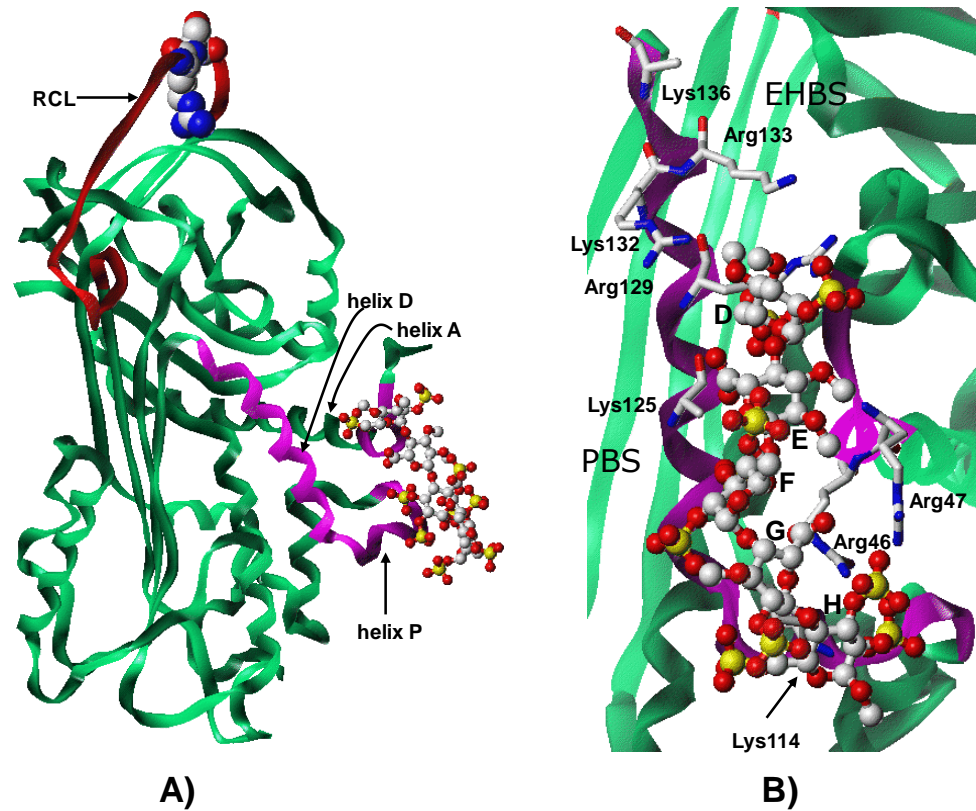
Figure 17. Stucture of heparin-complexed antithrombin. Part A shows the heparin-binding site and its position relative to the reactive center loop (RCL). Part B shows a zoomed-in version of the heparin-binding site along with critically interacting aminoacids. This figure is adapted from reference 167.

contribute ~50%, 25-33% and 28-35% of the total binding energy, respectively.[190-192] Other residues that play a minor role include Lys11, Arg13, Arg24, Arg46, Arg47 and Trp49.[193-195] Each of these residues contribute some 5-20 % of the total binding energy. In addition to interacting with the PBS, full-length heparin chains also interacts with residues Arg132, Lys133 and Lys136 located at the C-terminal end of helix.[196] The domain formed by these residues is called the extended heparin-binding site (EHBS). This interaction is supported by the fact that mutation of Arg132 and Lys133 selectively impairs the binding of heparin but not H5.

Although heparins and fondaparinux are effective at treating thrombotic disorders, their adverse side effect profiles make them less desirable anticoagulant medications. We reasoned that an alternative approach with nonsaccharide activators of antithrombin, molecules that are completely devoid of the heparin's saccharide scaffold, would be desirable. In our first attempt, we designed non-saccharide sulfated flavonoids (Figure 7) as mimics of trisaccharide DEF based on hydropathic interaction (HINT) analyses.[170,171] These sulfated flavonoids possess a structure dramatically different from heparin and were found to bind antithrombin with an affinity comparable to DEF and accelerated factor Xa inhibition approximately 10-fold. Competitive binding experiments suggested that the sulfated flavonoids by-pass the targeted PBS in AT and bind in the adjoining EHBS.[171] We hypothesized that an incremental increase in the size of the flavonoid skeleton may favor binding to the PBS. Based on this premise, a tetrahydroisoquinoline derivative, IAS5, was designed and evaluated along with a few analogs.

4.2. Rationale for the design of tetrahydroisoquinline derivative, IAS5

To improve on the antithrombin activation potential of these organic activators, a tetrahydroisoquinoline-based bicyclic-unicyclic sulfated activator IAS$_5$ was designed. Figure 18 shows the pharmacophore-based rational design of tetrahydroisoquinoline activators. Structure-activity studies show that three sulfate groups, i.e., the 6-*O*-sulfate on residue D and 3-*O*-sulfate and 2-*N*-sulfate on residue F, and the 6-carboxylate group of residue E are critical for conformational activation of antithrombin, while the trisaccharide scaffold only serves to position these groups for optimal interaction.[26] Thus, the pharmacophore was extracted and the four critical groups were connected in three-dimensional space using a linear carbon linker to arrive at a first 'blueprint' of an activator. The blueprint was transformed into a potential organic activator by introducing rigidity and simplifying the structure for rapid synthesis. In this process, several scaffolds were modeled and their similarity with the pharmacophore assessed. IAS$_5$, the tetrahydroisoquinoline-based activator containing an acid functionality and five sulfate groups, was selected because it retained the three-dimensional configuration of the four critical groups (Figure 19). In addition, the 2-*N*-sulfate of residue D was also effectively mimicked by a sulfate group of the isoquinoline ring. Calculation of the root mean square deviation between the five corresponding groups gave a value of 2.0 Å (range 1.1 to 2.7 Å), which reduced to 1.6 Å when only the groups of the pharmacophore were considered. In this design, it is suggested that the bicyclic ring mimics the D and E rings, the unicyclic ring mimics ring F, and the carboxylate groups match each other. To assess the importance of selected groups, we decided to study ester analogs, IES$_5$ and IES$_4$, as well as an acid
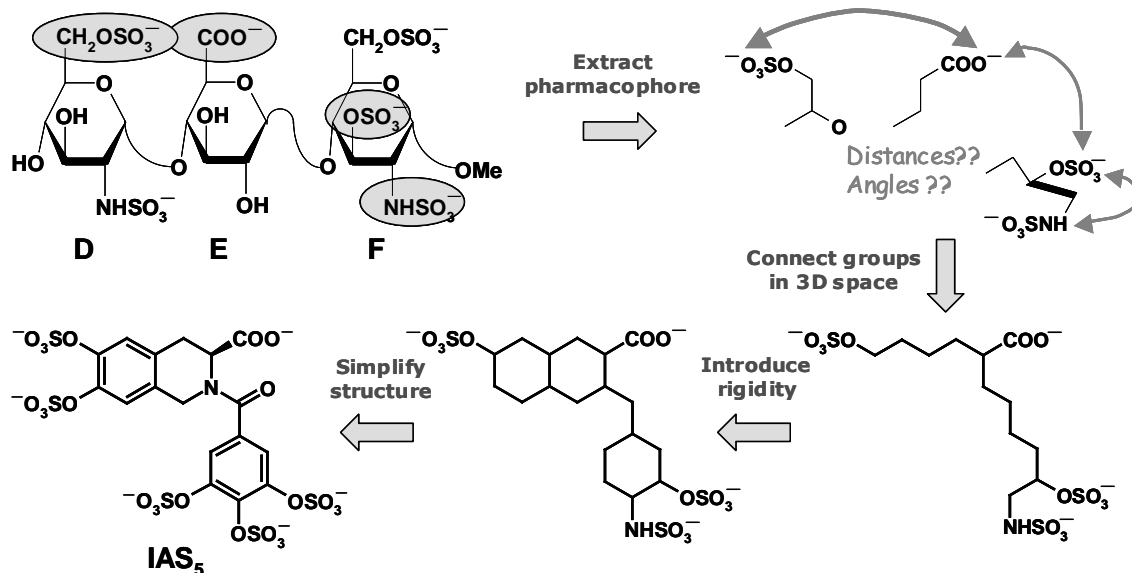
Figure 18. Rationale used in the design of tetrahydroisoquinoline-based organic activator IAS5. Four critical anionic groups (highlighted as filled ovals) of trisaccharide DEF formed the pharmacophore. Connecting the groups using a carbon framework followed by engineering of rigidity that matches their orientation gave rise to a 'blueprint', which was transformed into a synthetically plausible target, IAS_5

derivative with one less sulfate group, IAS_4 (Figure 20).

4.3. Results

4.3.1. Synthesis of IAS5 and its analogs

Nearly all small organic sulfates reported in the literature are mono- or di-sulfated molecules, typically prepared using sulfur trioxide complexes with amines in a highly polar solvent (see section 3.1 for a discussion). Yet, as the number of –OH groups increase
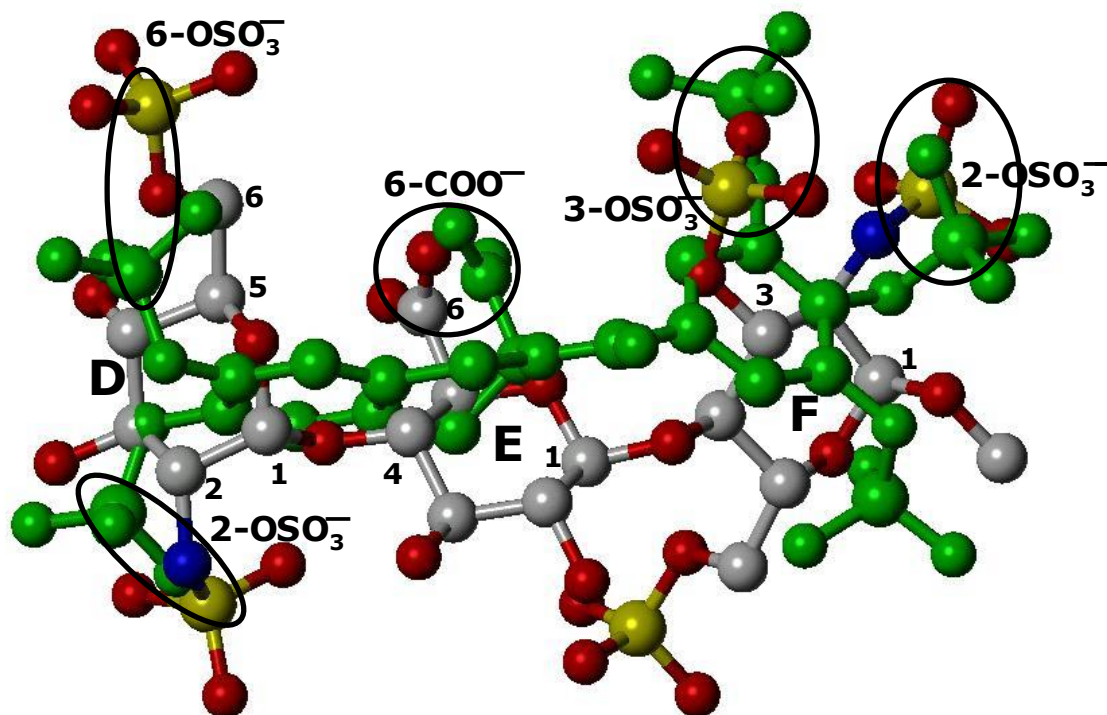
Figure 19. Overlay of DEF and IAS5 showing superposition of five anionic groups. The three critical sulfate groups – 6-OSO$_3^-$ on residue D and 3-OSO$_3^-$ and 2-OSO$_3^-$ on residue F – as well as the 6-COO$^-$ group of residue E overlay well on three sulfate and one carboxylate groups of IAS$_5$ in three dimensional space. In addition, the 2-OSO$_3^-$ group of residue D matches one of the remaining two sulfates on IAS$_5$. The overall RMSD for the five anions was 2.0 Å. See text for details.

on a small scaffold, sulfation becomes progressively more difficult because of anion crowding. We recently developed a rapid and high-yielding microwave-based synthesis of polysulfated organic molecules using trimethylamine– or pyridine– sulfur trioxide complexes in the presence of excess amine in anhydrous acetonitrile at $100 - 120\ {}^{O}C.$[4]

	X	Y
IAS5	SO_3^-	COO^-
IAS4	H	COO^-
IES5	SO_3^-	COOEt
IES4	H	COOEt
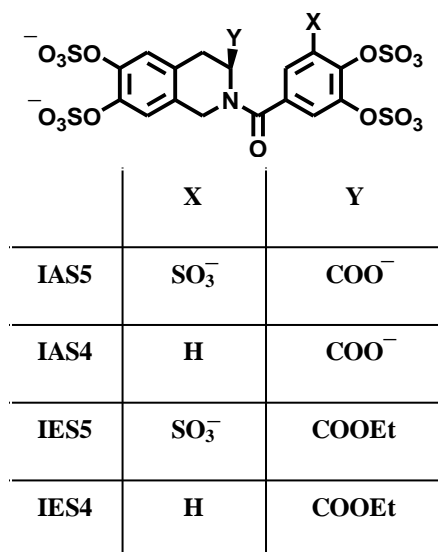
Figure 20. Structures of IAS5 and its analogs that vary in number of sulfate groups and/or presence/absence of free carboxylate group

Thus, ester analogs **IES₄ (4s)** and **IES₅** were prepared in four steps from commercially available starting materials (see Appendix A for details on the synthesis). However, the microwave-assisted sulfation failed to yield any sulfated products when the substrate was the polyphenolic molecule containing an acid group (Figure 21A). It is possible that the presence of naked carboxylic acid functionality under microwave conditions induces degradation of organic sulfates resulting in multiple products. To circumvent this problem, the carboxylic acid ester in IES_4 and IES_5 was hydrolyzed in excellent yields by simply stirring with KBuO-t / H_2O (2:1) in anhydrous DMSO at RT (Figure 21B).[197] It is worthwhile to mention that this hydrolysis fails miserably when standard saponification

A)

R = H/OH

R = H/OSO₃Na

B)

IES4/4s: $R_1 = R_2 = OSO_3Na$, $R_3 = H$
IES5: $R_1 = R_2 = R_3 = OSO_3Na$

IAS4: $R_1 = R_2 = OSO_3Na$, $R_3 = H$
IAS5: $R_1 = R_2 = R_3 = OSO_3Na$

Figure 21. Synthesis of acid derivatives **IAS4** and **IAS5** by selective cleavage of

corresponding ethyl esters (**B**). Direct sulfation of the precursor carboxylic acid results in

an intractable mixture of products (**A**).

conditions (NaOH/water) are used.

4.3.2. Equilibrium Dissociation Constant of Antithrombin Interaction with Sulfated Isoquinoline-based Activators

Previously we have used fluorescence spectroscopy for measuring the affinity of

ligands for antithrombin. For saccharide ligands, e.g., DEFGH and DEF, the ~30%

increase in intrinsic tryptophan fluorescence affords a convenient signal for the

determination of the K_D of the interaction,[27,168] while for our first generation sulfated

favonoids, we used an external probe, TNS, to determine binding affinity.[170,171] In this

study, we found that interaction of our designed organic ligands with antithrombin resulted

in a decrease in intrinsic tryptophan fluorescence that reached a plateau at high ligand concentrations (Figure 22). An equivalent limiting decrease of ~100% was obtained for IAS_5, IES_4 and IES_5, which could be fitted with the standard quadratic binding equation I (see section 4.5) to obtain the K_D of interaction at pH 7.4, I 0.15, 25 ^{O}C. K_D values of 320, 330 and 805 μM were measured for IAS_5, IES_4 and IES_5, respectively, corresponding to a similar free energy of binding between 4.2 and 4.8 kcal/mol (Table 4). For IAS_4, the fluorescence titration could not be made to reach an endpoint and, hence, a significantly weaker affinity (K_D > 1mM) is estimated. The affinity of reference trisaccharide DEF for antithrombin under similar conditions (pH 7.4, I 0.15, 25 ^{O}C) has not been reported, however a K_D value of 66 $\pm$ 4 μM has been measured in the absence of any added salt (pH 7.4, I 0.05, 25 ^{O}C).[168] In comparison, the affinities of IAS_5, IES_4 and IES_5 under these conditions were found to be 37, 50, and 130 μM (data not shown), respectively, suggesting that the designed activators compare favorably with trisaccharide DEF in antithrombin affinity. Likewise, the affinities of the first generation, flavonoid-based scaffolds under pH 7.4, I 0.15, 25 ^{O}C conditions were found to be approximately 130 $\pm$ 20 μM corresponding to a ΔG^O of 5.3 kcal/mol,[171] approximately 0.5 to 1 kcal/mol better than the 2nd generation agents.

4.3.3. Competitive Binding of Pentasaccharide DEFGH to Antithrombin in the Presence of IAS_5

To test whether the 2nd generation organic activators bind in the PBS, we studied the affinity of pentasaccharide DEFGH for antithrombin in the presence of varying
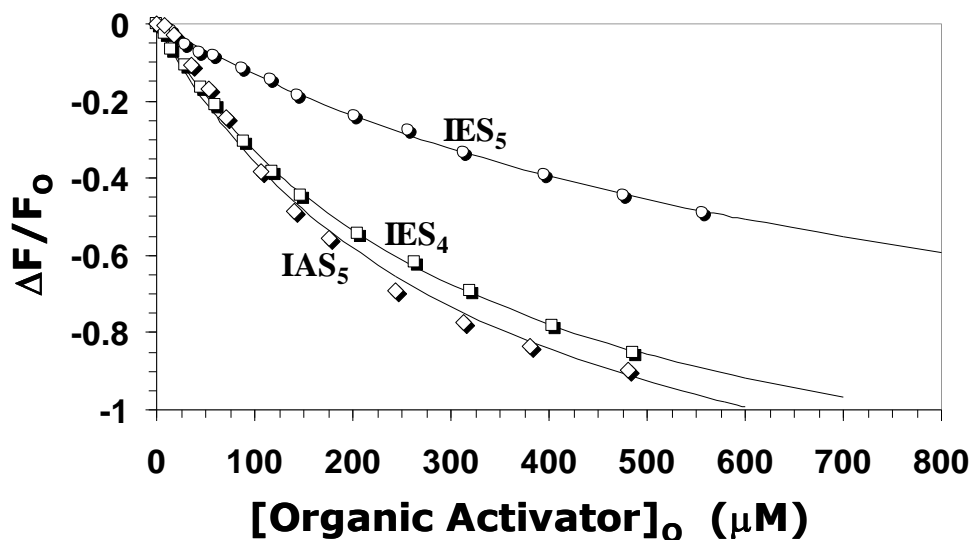
Figure 22. Fluorescence-based measurement of the equilibrium dissociation constant of organic activator – antithrombin complex at pH 7.4, I 0.15, 25 OC. Interaction of organic activators with antithrombin resulted in a saturable decrease in intrinsic tryptophans fluorescence at 340 nm (λ_{EX} = 290 nm), which was fitted to the quadratic binding equation I (section 4.5)to calculate the observed K_D. Solid lines represent the non-linear

Table 4. Thermodynamic and kinetic parameters for the interaction of the organic activators with plasma antithrombin at pH 7.4, I 0.15, 25 ^{O}C.

	$K_{D,OBS}$ $(\mu M)^a$	ΔF_{MAX} (%)	ΔG^O (kcal/mol)	k_{ACT} $(M^{-1}s^{-1})$	**Acceleration**
IAS$_5$	320 ± 10^c	-1.3 ± 0.2	4.8 ± 0.2	$69,780\pm580$	30
IES$_5$	805 ± 70	-1.2 ± 0.2	4.2 ± 0.4	6160 ± 240	2.7
IES$_4$	330 ± 20	-1.4 ± 0.2	4.8 ± 0.3	10860 ± 480	4.7
IAS$_4$	>1000	na	na	$5,370$	2.3

concentrations of a representative activator, IAS_5. The competitive binding experiment is made particularly easy because the two ligands induce opposite changes in intrinsic antithrombin fluorescence. While DEFGH induces ~30 – 35 % increase in intrinsic protein fluorescence at 340 nm,[27,168] IAS_5 induces ~100% decrease (above). Thus, if IAS_5 and DEFGH compete for the same binding site, it would be reasonable to expect that the affinity of DEFGH for antithrombin decreases in a manner predicted by the Dixon-Webb relationship.[*]

Figure 23 shows the observed affinity (K_D) and maximal fluorescence change (ΔF_{MAX}) in antithrombin – DEFGH titrations in the presence of increasing concentrations of IAS_5. The K_D value remained steady between 50±5 nM (0 μM IAS_5) and 31±7 nM (264 μM IAS_5) suggesting that the presence of the organic activator does not affect the interaction of DEFGH with antithrombin. Likewise, the ΔF_{MAX} value remained essentially constant between 37 (17 μM IAS_5) and 31 % (67 μM IAS_5), further supporting the independence of the two ligands in interacting with antithrombin. These results suggest that IAS_5 does not bind in the PBS. Alternatively, the results suggest that the 2nd generation activators may bind in the EHBS in a manner similar to the first generation agents.[171]

4.4. Discussion

Overall, the results show that non-saccharide IAS_5, designed on the basis of mimicking the trisaccharide pharmacophore activates antithrombin nearly 30-fold for the

---

[*] Dixon-Webb relationship states that for competing ligands L1 and L2, the affinity of L2 in the presence of L1 ($K_{L2,L1\neq0}$) would be related to that in its absence ($K_{L2,L1=0}$) through the relationship $K_{L2,L1\neq0} = K_{L2,L1=0}\times[1+[L1]_O/K_{D,L1}]$.
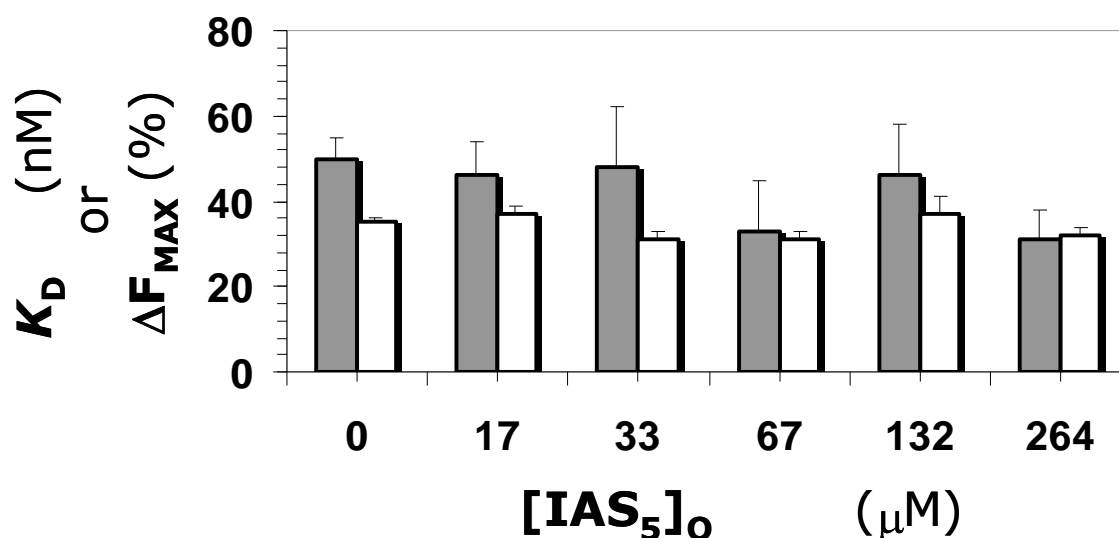
Figure 23. Competitive binding of pentasaccharide H5 to antithrombin in the presence of organic activator IAS5. Equilibrium dissociation constants (K_D, shaded bars) and maximal change in fluorescence (ΔF_{MAX}, open bars) of fluorescence titrations of DEFGH binding to antithrombin at pH 7.4, I 0.15, 25 ^{O}C in the presence of varying concentrations of IAS_5 (0 to 264 μM) are shown. Error bars represent ±1 S. E

accelerated inhibition of factor Xa. This represents a 2-3-fold improvement in acceleration in comparison to that achieved with the first generation flavonoid-based activators. Despite the advance, IAS_5 is approximately 10-fold weaker than trisaccharide DEF (or pentasaccharide DEFGH) suggesting that major structural improvements in the organic scaffold are necessary.

The key reason for the difference in the antithrombin activation potential between IAS_5 and DEF (or DEFGH) is the difference in their site of binding. Whereas IAS_5 does

not engage the pentasaccharide-binding site (PBS), DEF (and DEFGH) is known to bind in the PBS. This implies that IAS_5 does engage the three members of the electropositive triad, Lys114, Lys125 and Arg129, which is known to play a crucial functional role in the mechanism of heparin binding and conformational activation of antithrombin.[167,198] Thus, IAS5 is similar to sulfated flavonoids with respect to site of binding.

Despite these results, it is interesting that IAS5 is able to significantly enhance the inhibition of factor Xa. A plausible explanation is that IAS_5 binds to the activated form of antithrombin that exists in pre-equilibrium with the native form, thereby altering the equilibrium more in favor of the activated state. This mechanism of antithrombin activation has been noted earlier with smaller oligosaccharides that possess relatively weak affinity for the serpin.[168]

4.5. Experimental section

Synthesis and Characterization of Organic Activators* — IES_4 and IES_5 were synthesized from corresponding methoxy-protected intermediates in two steps, namely demethylation and sulfation (see Appendix A). The carboxylate derivatives IAS_4 and IAS_5 were synthesized from IES_4 and IES_5, respectively. Briefly, each ester (0.03 – 0.08 mmol) was stirred with potassium *t*-butoxide and H_2O (2:1 molar equivalent, 3 equiv. KBuO-*t*) in DMSO (0.1 M) under nitrogen at room temperature until the starting material was consumed (HPLC analysis, 1 – 3 h),[16] after which sodium dibasic phosphate (10 equiv) in H_2O (2 mL) was added to the reaction mixture. Following 15 min of stirring, the reaction

---

* Detailed synthetic procedures, schemes and capillary electropherograms may be found in Appendix A

mixture was loaded onto a Sephadex G10 column (160 cm) and chromatographed using water as an eluent. Fractions were combined, concentrated under vacuum and lyophilized to obtain the sodium salt of the carboxylate as a white solid. Capillary electrophoresis of both products using a fused silica capillary in 20 mM sodium phosphate buffer, pH 2.3, at 10 kV showed a single peak with greater than 95% purity (see Appendix A). **IAS$_4$**: ^{1}H-NMR (400 MHz, [2]H$_2$O): δ 6.99-7.51 (m, 5H), 4.72 – 4.85 (m, 2H), 4.42 – 4.52 (m, 1H), 3.01-3.16 (m, 2H); ESI (-ve) m/z calcd for C$_{17}$H$_{10}$NNa$_5$O$_{19}$S$_4$ [(M-Na)$^-$] 751.83, found 751.91; **IAS$_5$**: ^{1}H NMR (400 MHz, [2]H$_2$O) δ 7.05 – 7.46 (m, 5H), 4.75-4.99 (m, 2H), 4.47-4.53 (m, 1H), 3.03–3.11 (m, 2 H); ESI (-ve) m/z calcd for C$_{17}$H$_9$NNa$_6$O$_{23}$S$_5$ [(M+H)$^+$] 869.77, found 869.83.

Proteins and Chemicals — Antithrombin and factor Xa (human forms) was purchased from Haematologic Technologies (Essex Junction, VT) and used as received. Molar concentrations of antithrombin were calculated from absorbance measurements at 280 nm using a ε$_{MAX}$ of 37,700 M^{-1}cm^{-1}. Antithrombin was stored in 20 mM sodium phosphate buffer, pH 7.4, containing 100 mM NaCl, 0.1 mM EDTA and 0.1% (w/v) PEG8000 at -78 ^{O}C, while factor Xa was stored in 5 mM MES buffer, pH 6.0, containing 25 mM NaCl at -78 ^{O}C until use. Factor Xa substrate Spectrozyme FXa was obtained from American Diagnostics, Greenwich, CT. Trimethylamine-sulfur trioxide complex was purchased from Alfa-Aesar (Ward Hill, MA). All other reagents/chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

Experimental Conditions — Antithrombin interaction and activation studies were performed at 25 ^{O}C and in 20 mM sodium phosphate buffer, containing 100 mM NaCl, 0.1 mM EDTA and 0.1% (w/v) PEG 8000, adjusted to 7.4. The ionic strength of this buffer is 0.135 and the buffer is labeled as pH 7.4, *I* 0.15, 25 ^{O}C buffer.

Fluorescence Spectroscopy and Equilibrium Binding Studies — Fluorescence experiments were performed using a QM4 fluorometer (Photon Technology International, Birmingham, NJ) in pH 7.4, *I* 0.15, 25 ^{O}C buffer. Equilibrium dissociation constants (K_D) for the interaction of organic activators with plasma antithrombin were determined by titrating the activator into a solution of plasma antithrombin and monitoring the decrease in the fluorescence at 340 nm (λ_{EX} = 290 nm). The slit widths on the excitation and emission side were 1 mm and 2 mm, respectively. The decrease in fluorescence signal was fit to the quadratic equilibrium binding equation I to obtain the K_D of interaction, wherein ΔF represents the change in fluorescence following each addition of the activator ($[ACT]_O$) from the initial fluorescence F_O and ΔF_{MAX} represents the maximal change in fluorescence observed on saturation of antithrombin ($[AT]_O$).

$$\frac{\Delta F}{F_O} = \frac{\Delta F_{MAX}}{F_O} \times \frac{([AT]_O + [ACT]_O + K_D) - \sqrt{([AT]_O + [ACT]_O + K_D)^2 - 4*[AT]_O[ACT]_O}}{2*[AT]_O}$$
……**I**

Competition between DEFGH and IAS$_5$ binding to antithrombin was studied in pH 7.4, *I* 0.15 buffer at 25 ^{O}C by monitoring the increase in fluorescence at 340 nm (λ_{EX} = 280 nm) as a function of the concentration of DEFGH. Several titrations were performed at

different concentrations of activator IAS$_5$ (0 to 264 μM) and the K_D and ΔF_{MAX} values were calculated using equation 1.

Factor Xa Inhibition Studies — The kinetics of inhibition of factor Xa by antithrombin in the presence of organic activators under pseudo-first order conditions was measured spectrophotometrically in a manner similar to our earlier work.[10,12] A fixed 10 nM concentration of factor Xa was incubated with fixed concentrations of plasma antithrombin (0.1 to 0.5 μM) and the sulfated activator (0 to 100 μM) in pH 7.4, *I* 0.15 buffer at 25 ^{O}C. At regular time intervals, an aliquot of the inhibition reaction (100 μL) was quenched with 900 μL of 100 μM Spectrozyme FXa in pH 7.4, *I* 0.15 buffer and the initial rate of substrate hydrolysis was measured from the increase in absorbance at 405 nm. The exponential decrease in the initial rate of substrate hydrolysis as a function of time was used to determine the observed pseudo-first order rate constant of factor Xa inhibition (k_{OBS}). A plot of k_{OBS} values measured as a function of different concentrations of an organic activator could be described by equation II, in which k_{UNCAT} is the second-order rate constant of factor Xa inhibition by antithrombin alone, i.e., 2300 M^{-1}s^{-1} at pH 7.4, I 0.15, 25 ^{O}C,[10] and k_{ACT} is the second-order rate constant of factor Xa inhibition by antithrombin – organic activator complex.

$$\frac{k_{OBS}}{[AT]_O} = k_{UNCAT} + k_{ACT} \times \frac{[ACT]_O}{[AT]_O + K_D} \text{ .....}\mathbf{II}$$

COMBINATORIAL VIRTUAL SCREENING OF HEPARIN & HEPARAN SULFATE

5.1. Introduction

Glycosaminoglycans heparin and heparan sulfate play diverse roles in a number of physiological and pathological processes including coagulation, angiogenesis, immune response and viral infection.[*] These functions originate from their interaction with numerous proteins, which include serpins, antithrombin and heparin cofactor II, coagulation proteinases, fibroblast growth factors and their receptors, cytokines and viral cell envelope glycoproteins. It is commonly assumed that these roles arise from an optimal combination of specificity and affinity, yet, except in a few cases, the specificity of heparin and heparan sulfate (HS) interactions remain ill-defined and poorly understood.[**]

A major reason for the ill-defined structure-activity relationships in these heparin/HS interactions is the phenomenal structural diversity of heparin/HS glycosaminoglycans (GAGs), and its concomitant difficulties, especially synthesis,

---

[*] See section 1.1 for a discussion of the physiological roles of GAGs
[**] For a discussion of specific GAG sequences, see section 1.4

67

purification, and structure identification.[*] Both heparin and heparan sulfate are complex, highly anionic polysaccharides composed of alternating 1→4-linked glucosamine and uronic acid residues, which are variously modified through sulfation, acetylation and epimerization.[199] These modifications can produce 48 different building blocks, or disaccharides, of which 23 have been found to date. Further, the iduronic acid residue (IdoAp) can exist in multiple conformations, especially $^1C_4$ and ^2S_O for internal locations, that can inter-convert relatively easily.[111,125] Thus, combination of structural and conformational variability generates millions of sequences, of which few are expected to recognize a target protein. This expectation is further borne out by the growing evidence that the heparan sulfate biosynthetic apparatus has considerable specificity and organization.[**]

On the protein front, considerable effort has been made in trying to deduce consensus binding sequences that recognize heparin structures. Whereas Cardin and Weintraub suggested linear consensus sequences with specific repeat pattern to be important for heparin binding, others have suggested a spatial distance relationship.[200-202] Although it is clear that arginine and lysine residues lining protein surfaces dominate heparin-binding sites, their optimal 3D orientation that generates high specificity and affinity remains unclear. The situation is further compounded because not all arginine and lysine interactions with sulfate and carboxylate groups of heparin/HS are identical. While heparin interaction with antithrombin and basic fibroblast growth factor involves ~40 and

---

[*] See section 1.5 for an extensive discussion of the problems associated with defining GAG SAR
[**] Section 1.3 discusses the biosynthesis of heparin and HS

30% ionic binding energy,[168,203] respectively, that with thrombin involves ~80%.[204] Thus, identifying the precise heparin-binding site on proteins, although expected to be straightforward, has not been an easy task.

In the past decade fifteen heparin-protein co-crystal or NMR structures have become available. These include complexes with thrombin,[115] growth factor and growth factor receptors,[50,52,53] annexin V,[116] heparan sulfate 3-O-sulfotransferase,[117] RANTES[121] and antithrombin. Of these, antithrombin is the most studied protein with five co-crystal structures detailing the interaction of a five-residue sequence with the inhibitor in either a binary or a ternary complex.[118-120,122] Yet, these represent a small fraction of the vast number of physiologically active complexes. Further, except for antithrombin, the 'heparin' structure reported in these complexes is the fully sulfated saccharide (see Figure 3B), the most common structure found in heparin, which may not be the optimal sequence.

In view of the limited structural knowledge on GAG–protein interactions and the phenomenal structural diversity of HS, computational docking approaches represent a powerful means of assessing binding affinity and specificity. In fact, the computational literature is replete with numerous GAG studies, yet no approach has been devised that predicts high specificity sequences. Modeling GAGs is challenging because of their high negative charge density, which induces recognition of practically any collection of positively charged residues, and their surface–exposed, shallow binding sites on proteins.[50,52,53,115-117,121] With respect to antithrombin in particular, two computational attempts have been made to understand heparin binding with mixed results.[205] Both Grootenhuis and van Boeckel[206] and Bitomsky and Wade[205] attempted to derive the

heparin pentasaccharide binding site on antithrombin using molecular dynamics and docking, but their modeled geometry turned up to be significantly different from the co-crystal structure. Similar challenging results were noted with our small, sulfated molecules, which were rationally designed to recognize the pentasaccharide-binding site in antithrombin but were found to preferentially interact with the extended-heparin binding site.[170,171,173] Thus, we reasoned that a robust docking protocol for GAGs in general, and heparin/HS in particular, would be very useful for understanding specificity of their interactions.

This chapter describes our novel approach of predicting high specificity GAG sequences with the well-studied antithrombin–heparin interaction as a test case. Our approach is based on a two-filter strategy, with the first step involving an affinity filter followed by a geometry convergence filter using a genetic algorithm-based docking tool, GOLD[TM].[207] The two-filter strategy rapidly sorted a combinatorial virtual library of nearly 7,000 heparin hexasaccharides into specific and non-specific sequences,[*] thus suggesting its potential use for identifying 'needle(s) in a haystack'.

5.2. Results and Discussion

5.2.1. GOLD[TM] predicts the binding geometry of synthetic pentasaccharide H5$_{CRYS}$ to within 0.6 Å

---

[*] In this chapter and chapter 6, 'specificity' refers to the existence of a few, structurally unique HS sequences from a combinatorial library of all possible sequences that can recognize the protein binding site in a single, well-defined binding mode. Therefore, specific HS sequences were determined by performing multiple molecular docking experiments to assess the reproducibility of computed binding modes.

The ability to predict conformation of GAG in the receptor-bound state was performed using a rigid body docking procedure in GOLD^{TM}. The rigid body docking procedure reduces the search space significantly and allows an assessment of GOLD's fitness function, which is comprised of four energy terms – external and internal hydrogen bonding and van der Waals energy. Docking was evolved for 100,000 iterations in each run to ensure sampling of most of the conformational space available in the pre-defined binding site on antithrombin. Further, to ensure greater confidence in docked geometries, the process was repeated three times.

Successful docking is typically evaluated by RMS difference between the docked solution(s) and the X-ray geometry, if available. While a RMSD value of less than 1.0 Å would indicate near identity of the two geometries, values between 1. 0 and 3.0 Å have been utilized by many researchers for correctly docked geometries of drug-like small molecules (MW <500).[208-210] Considering that our penta- and hexasaccharides are significantly bigger (MW ~2,000) and the resolution of antithrombin–heparin–thrombin crystal structure was only 2.5 Å,[119] we employed a 2.5 Å RMSD criterion. When the synthetic pentasaccharide H5$_{CRYS}$ was docked onto antithrombin (Figure 24), the GA search pre-terminated in all three independent runs, each with a high GOLD score of greater than 100. All 6 solutions were identical with an inter-solution RMSD of less than 0.6 Å suggesting a strong preference for this geometry in the binding site (not shown). Comparison of these solutions with the H5$_{CRYS}$ structure present in the crystal suggests identical orientation and binding site. The maximal RMSD between the docked solutions and the crystal structure was also 0.6 Å. In addition, interaction at the atomic level for the
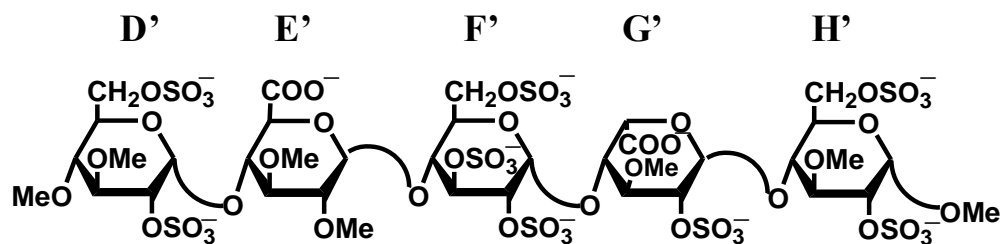
Figure 24. Structure of a H5 derivative, H5$_{CRYS}$, extracted from the crystal structure study of Li et al[119] and used in docking studies. Note that all N-sulfate and free hydroxyl groups of H5 are modified to O-sulfate and OMe groups, respectively.

docked and co-crystal complexes were found to be essentially identical, suggesting that the rigid-body docking procedure can reliably identify the binding geometry and the binding site of the most active pentasaccharide sequence.

5.2.2. Prediction of binding geometry of natural pentasaccharide sequence DEFGH with an "average GAG backbone" conformation to within 2.0 Å

X-ray fiber diffraction studies indicate that GAGs adopt a helical structure,[123] which has now been confirmed through several crystal structures.[50,52,53,115-119,121] For heparan sulfate-like GAGs (HL-GAGs), the helix has a 2_1 symmetry with inter-glycosidic torsion angles (ϕ_H [H_{UA}-C_{UA}-O-C_{GlcNp}] and ψ_H [C_{UA}-O-C_{GlcNp}-H_{GlcNp}]) constrained to generate the helix. These conclusions are borne out in a number of NMR and molecular dynamics studies on heparin[111-113] and heparin oligosaccharides,[109,110,114] which show that the ϕ_H/ψ_H angles vary within a relatively narrow range in solution. Thus, heparin appears

to maintain a 2-fold helix irrespective of its sequence and substitution patterns. Additionally, crystal structures of pentasaccharide-antithrombin[118-120,122] and hexasaccharide-FGF[53] complexes demonstrate that only small changes in ϕ_H/ψ_H angles occur on protein binding, a result also observed with NMR studies of pentasaccharides bound to AT.[109,114] Compiling these results indicates that ϕ_H and ψ_H values vary from 35° to 61° and 6° to 25°, respectively, for the UAp(1→4)GlcNp inter-glycosidic linkage, while ϕ_H varies between -68° and -9° and ψ_H varies between -60° and -1^O for GlcNp(1→4)UAp linkage. These variations represent a maximal change of only 19 to 59°, or less than $\pm30^O$, a relatively small deviation considering the large number of possibilities for a structurally diverse molecule. Thus, we reasoned that HL-GAGs, irrespective of their sequence and substitution pattern, could be simulated by an "average" structure, wherein the ϕ_H/ψ_H inter-glycosidic bond angles are held constant at the mean of known solution values. This implies that docking of any heparin/HS sequence with an "average backbone" model should reliably simulate the physiological GAG binding structure. To test our hypothesis, the natural sequence-specific pentasaccharide H5 (Figure 7) was modeled in which inter-glycosidic torsions were held at the 'average backbone' values, rather than set to those obtained from the crystal structure of H5$_{CRYS}$. Docking was performed in which the inter-glycosidic torsions and intra-ring conformations were held constant, while substituents present at the 2-, 3-, and 6- positions were allowed conformational flexibility. All three docking experiments terminated rapidly, indicating a strong preference for one
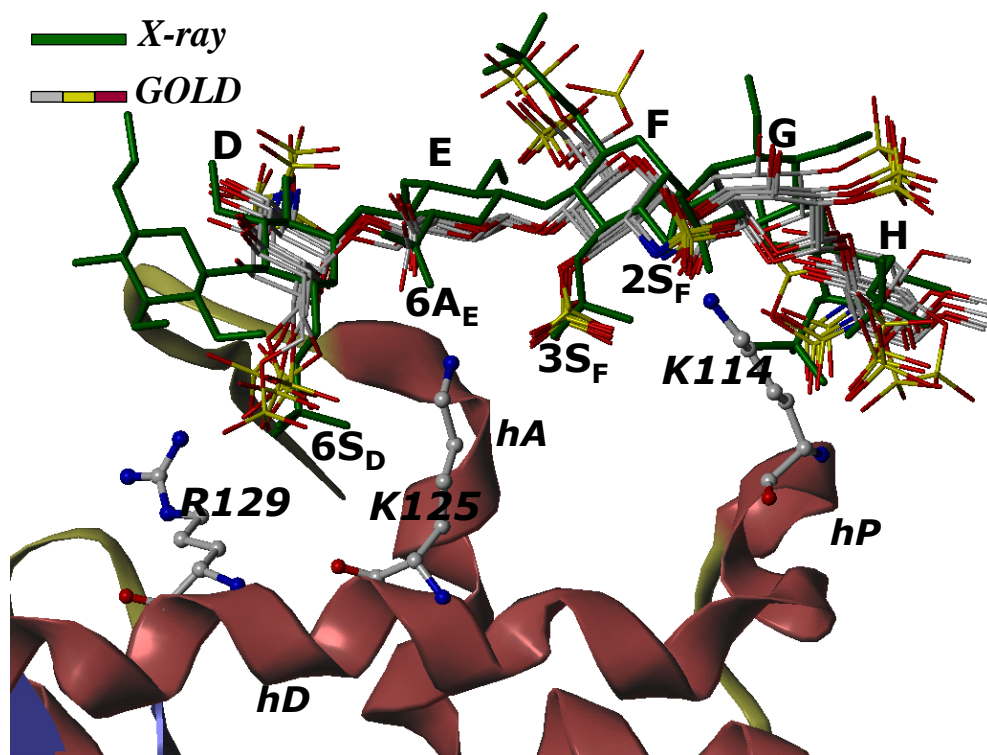
Figure 25. Comparison of GOLD^{TM} predicted binding geometry of natural

pentasaccharide H5 having 'average backbone' with that of H5$_{CRYS}$ determined in the

crystal structure. An overlay of 6 solutions from three independent docking runs shows

high consistency in the predicted binding geometry, which matches the crystal structure

geometry with an RMSD less than 2.5 Å. Structure in green is the crystal structure

geometry, while those in atom-type color (red, yellow, grey and blue) are 6 docking

solutions. Note the identical orientation of key groups, 2- and 3-OSO$_3^-$ of residue F (2S$_F$

and 3S$_F$), 6-COO$^-$ of residue E (6A$_E$) and 6-OSO$_3^-$ of residue D (6S$_D$). Helices A, D

and P of antithrombin (in ribbon diagram) are indicated as *hA*, *hD* and *hP*, while D, E,

F, G, and H labels correspond to residues of the pentasaccharide. K114, K125 and R129

are shown in ball and stick representation.

type of binding geometry. Figure 25 shows a comparison of the 6 GOLD solutions with the geometry present in the crystal structure. All 6 solutions predict an essentially identical binding geometry, an observation typically not found in such statistical docking studies. Small differences are found for the orientation of the terminal $-OSO_3^-$ and $-OCH_3$ groups, especially in the G and H residues. All 6 solutions were within 2.5 Å RMSD (Figure 25) of each other and comparison of these structures with $H5_{CRYS}$ present in the crystal structure indicates nearly identical orientation, conformational state and interactions (Figure 25).Analysis of the GOLD solutions at the atomic level facilitates identification of groups important for specificity of interaction. For example, the average RMSDs for the sulfate groups at positions 2 and 3 of residue F were found to be 0.2±0.1 and 0.3±0.2 Å indicating a virtually invariant conformation in all solutions. Likewise, the average RMSDs for sulfates at 6 and 2 positions of residues D and H, respectively, were found to be 1.1±0.5 and 1.0±0.8 Å suggesting reduced tendency for alternative rotational states. In contrast, the average RMSDs of sulfates at the 6-, 2- and 6-positions of residues F, D and H were 2.4±1.2, 2.1±1.2 and 1.6±0.9 Å, respectively, indicating greater conformational variability. Our docking results suggest that four sulfate groups, at positions 6-, 2- and 3-, and 2- of residues D, F and H, respectively, have minimal conformational variability in the binding site. Thus, these sulfate groups organized in a specific three-dimensional orientation afford the specificity of interaction, a conclusion proposed earlier on the basis of a large number of structure-activity studies (see Figure 7).[146,167,168,211,212]

5.2.3. Prediction of binding geometry with a flexible H5 fails to give reasonable solutions

The above docking procedure utilized constrained intra- and inter-glycosidic conformations to maintain solution-state structure in the protein-bound state. To test whether full conformational flexibility during docking would result in better predictability, docking of H5 was performed under conditions wherein the HL-GAG backbone and saccharide conformations were unrestrained. Striking differences were noted in these docking studies. None of the three docking experiments converged suggesting a lack of preference for a defined binding geometry (not shown). In addition, the final six solutions display completely random profiles. Further, the pyranose rings in the docked solutions were found to exhibit conformations not found in solution. To test whether partial flexibility is tolerated, H5 was docked with unrestrained inter-glycosidic torsions, but constrained intra-ring torsions matching geometries in their bound state. The RMSD matrix showed an inconsistent set of 6 solutions corresponding to several different binding geometries (not shown). Only 1 of these geometries was within 2.5 Å RMSD of H5$_{CRYS}$. These experiments suggest that determining binding geometries with a fully flexible ligand, though highly desirable, is difficult. These results demonstrate that the imposition of an average backbone constraint significantly increases the chances of arriving at the experimentally-derived binding geometry by reducing the conformational search space.
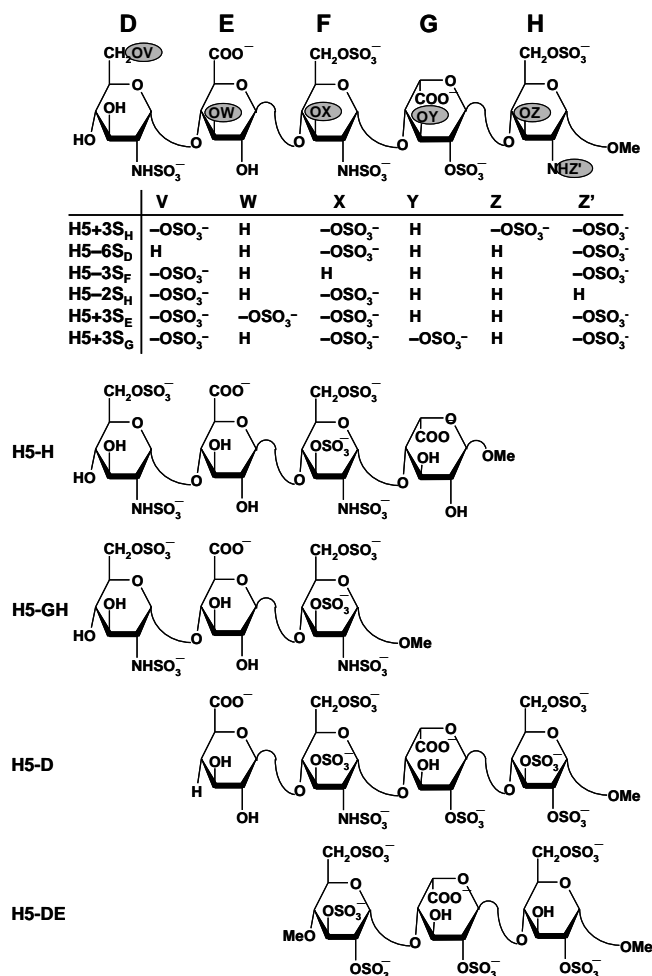
Figure 26. Structures of H5 and its truncated variants studied for docking . D, E, F, G, and H refer to residue labels. Variations in structure (V, W, X, Y, Z, and Z') to give variants with either one more sulfate group (H5+3S$_H$, H5+3S$_E$ and H5+3S$_G$) or one less sulfate group (H5-6S$_D$, H5-3S$_F$, H5-2S$_H$) are highlighted as filled ovals. Truncated pentasaccharides H5-H, H5-GH, H5-D, and H5-DE refer to one or two residue deletion variants from either end.

5.2.4. The docking protocol sorts pentasaccharides based on specificity of interaction with antithrombin.

The natural pentasaccharide sequence H5 is a rare sequence in heparin in which the EF unit (Figure 7) with the presence of the 3-O-sulfate group is both unique and interesting. Studies with a large number of H5 derivatives have led to the understanding that other sulfate groups present in the pentasaccharide sequence, e.g., the N-sulfate of D and 6-O-sulfate of F, are not important.[26] Interestingly, certain analogs with sulfation level higher than H5, e.g., H5+3S$_E$ and H5+3S$_G$ (Figure 26), were less active than H5 in sharp contrast to the expectation that greater sulfation level, especially 3-O-sulfation, would induce higher antithrombin-affinity.[26] At the same time, studies by Desai et. al. suggested that 3-O-sulfation of residue H was not detrimental to biological activity.[168,213]

An obvious explanation for the changes in antithrombin affinity is the loss or gain of favorable or unfavorable interactions. Yet, this explanation suggests that individual interactions affect the changes independently. To gain a more fundamental understanding, we studied the docking of several H5 variants including H5−3S$_F$, H5−6S$_D$, and H5−2S$_H$ (Figure 26), each lacking a sulfate at position 3 in F, position 6 in D, and position 2 in H, respectively, onto antithrombin. In addition, H5+3S$_E$ and H5+3S$_G$, described above, and a heparin hexasaccharide with common repeat sequence GH (IdoAp2S−GlcNp2S6S) were also studied. Finally, a H5 sequence with IdoAp residue locked in the unfavorable $^1C_4$ conformation (H5/G*) was studied to understand the importance of conformational flexibility in this residue.
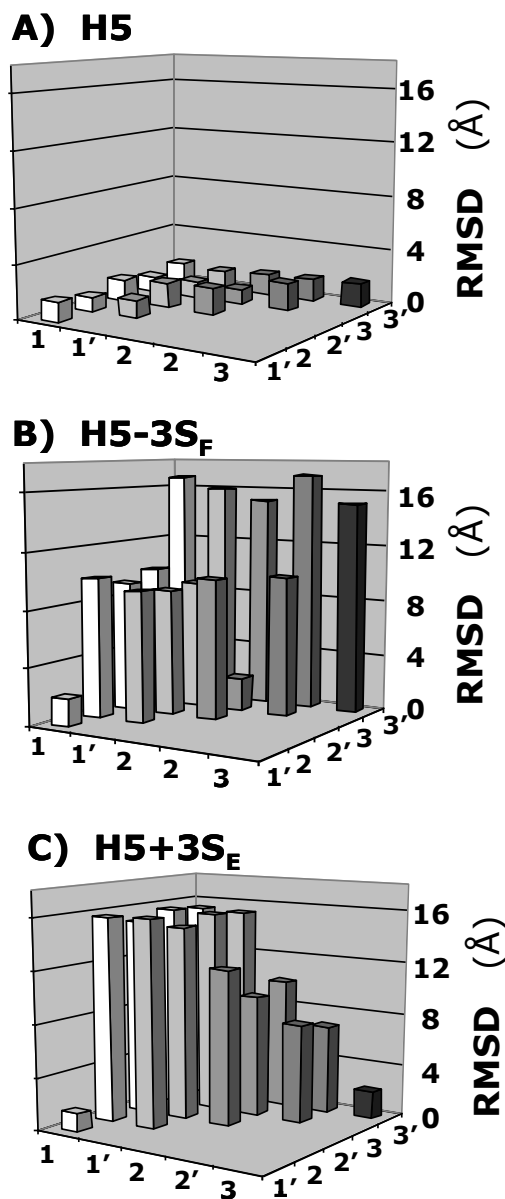
Figure 27. Specificity of GAG sequence for antithrombin. Variation in the RMSD among

six solutions (1, 1', 2, 2', 3, and 3') obtained from three independent docking experiments

of heparin pentasaccharide H5 variants, containing either one additional sulfate group at 3-

position of residue E or lacking a sulfate at the 3-position of residue F, binding to

antithrombin.

These H5 analogs were prepared *in silico,* minimized at the average GAG ϕ_H/ψ_H values, and docked using the protocol described above for H5. Figure 5 shows the RMSD among the solutions obtained in three docking experiments for each of H5 variant. Whereas for H5, all 6 solutions are essentially identical with RMSD less than 2.5 Å (Figure 27A), none of the variants display this behavior. For example, H5−3S$_F$, lacking the critical 3-*O*-sulfate group, displays only two RMSDs within 2.5 Å (Figure 27B), while H5+3S$_E$, possessing an additional 3-*O*-sulfate in residue E, displays only one (Figure 27C). Likewise, the number of docked solutions that are within 2.5 Å RMSD of each other is 8, 7, and 4 for H5−2S$_H$, H5−6S$_D$, and (GH)$_3$, respectively, out of the possible 15, while that for H5+3S$_G$ and H5/G* are 6 and 1, respectively (not shown). Thus, whereas 100% of solutions are found within 2.5 Å of each other for H5, only 53%, 47%, and 13% are found for H5−2S$_H$, H5−6S$_D$, and H5−3S$_F$, respectively. For H5+3S$_E$ and H5+3S$_G$, variants with additional 3-*O*-sulfation, this percentage is 13 and 40%, while for $^1C_4$ conformationally locked H5 (H5/G*) it is only 6.7%.

   The above results indicate that self-consistency of docking geometries is sensitive to sulfate group distribution and pentasaccharide topology. Alternatively, the loss or gain of sulfate groups or $^1C_4$ iduronate conformation gives rise to alternate binding modes significantly differing from each other and from that of the natural high-affinity sequence H5. A closer inspection of the binding geometries suggests that many of these variant sequences favor a completely different pattern of interactions. These altered preferences, especially a multitude of them, are likely to be non-productive generating significant non-
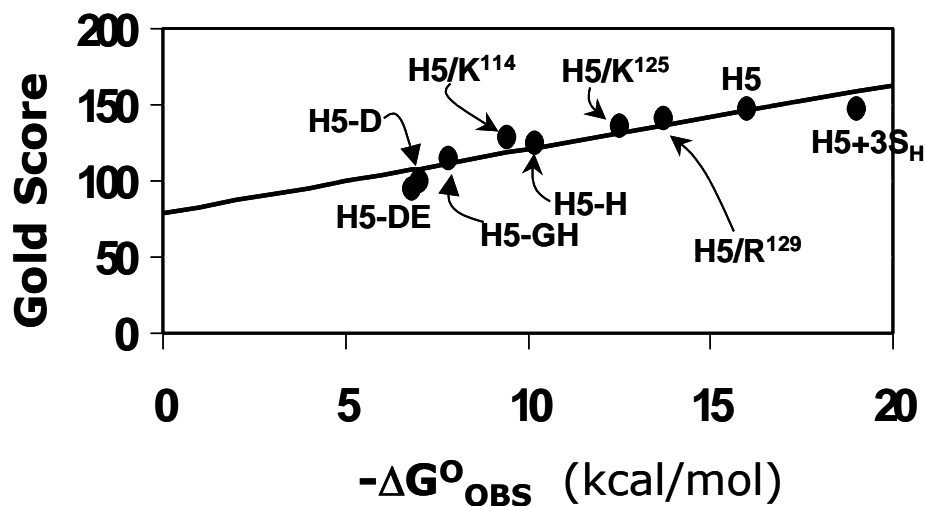
Figure 28. Correlation of GOLD score with antithrombin binding affinity.

Modified GOLD score (See experimental section) determined following

100,000 iterations and multiple docking runs linearly increases with the

observed antithrombin binding affinity under physiological conditions. Only

those antithrombin complexes for which binding affinities were measured

under essentially equivalent conditions are used for analysis.

specific binding, thus resulting in loss of binding affinity. Thus, these docking experiments indicate that the 'self-consistency' of the RMSD matrix of top-ranked solutions is a sensitive measure of GAG specificity.

5.2.5. GOLD scores correlate with antithrombin affinity

Implicit in our docking protocol is the assumption that GOLD scores correlate with protein binding affinity. To establish that this is indeed true, we studied the docking of pentasaccharide variants with antithrombin for which the binding affinity had been measured under standardized conditions.[168,190-192,213] These include H5, H5+3S_H, H5–GH, H5–H, H5–D, and H5–DE binding to wild-type antithrombin[2,26,57] and H5 binding to R129H, K125M and K114Q antithrombin mutants.[58-60] These 9 complexes represent a range of ~12 kcal/mol in free energy of binding, a wide range spanning very strong to very weak interactions. Docking was performed in a manner identical to that for native pentasaccharide H5 and the top solution analyzed for binding energy correlation.

Figure 28 shows a profile of modified GOLD score *versus* the observed free energy of binding under physiological conditions. For the 9 complexes studied, the modified GOLD score increased linearly from ~95 to ~150 giving a slope of 4.2 GOLD score per kcal/mol and an intercept of 78.9. The reasonable linear correlation between modified GOLD score and ΔG^O_{OBS} validates the docking protocol. Yet, the slope of 4.2 GOLD score per kcal/mol reflects a less-than-optimal sensitivity to structural change, especially considering that multiple docking geometries were not considered in the analysis. In general, predicting protein-binding affinities of GAGs has been difficult because of the

exposed nature of the binding site and the unquantified interaction parameters of sulfate groups. In addition, knowledge-based or empirical scoring functions do not take into account entropic contributions to ΔG^O, which arise from the release of Na^+ and H_2O species from antithrombin and GAG surfaces following saccharide binding and are the dominant factors in the stabilization of the complex. Finally, the scoring function cannot compute the energy required for conformational flips of iduronate residues as well as the energy consumed to induce an allosteric, two-step binding interaction necessary for antithrombin.[168,213] In light of these unknowns, the correlation observed in Figure 28 is reasonably good, especially considering that the affinity varies from the high μM to the low nM.

5.2.6. Identification of high-affinity, high specificity sequences through a combinatorial virtual screening approach

Although a large number of proteins interact with HL-GAGs in our body, little information is available on sequences that recognize these proteins. This is also the status for other GAGs, such as dermatan sulfate, chondroitin sulfate and keratan sulfate. In addition, knowledge regarding the specificity of these interactions is completely lacking. We reasoned that our protocol, which combines two functions – affinity in the form of GOLD score and specificity in the form of consistency of docking – represents a powerful tool to deduce high-affinity GAG sequences that bind proteins with high specificity.

To test this hypothesis, we generated a combinatorial library of heparin hexasaccharides for docking onto antithrombin. The library was generated from a set of 19

disaccharides (Figure 4), which included all monosaccharides found in nature, except for the rare free glucosamines.[104] The library considered the conformational flexibility of IdoAp residues in an explicit manner through inclusion of both $^1C_4$ and ^2S_O conformations. Thus, the combinatorial library consisted of 6,859 unique heparin hexasaccharides built with the 'average backbone' geometry in an automated manner.

A biphasic docking protocol was used for screening this library. The first phase evaluated the binding of each hexasaccharide onto antithrombin using 10,000 iterations, instead of 100,000, followed by GOLD score screening. The histogram of GOLD score frequency obtained from this analysis suggests an approximately Gaussian distribution of sequences centered between 40 and 55 GOLD score with the highest at 115 (Figure 29). A majority of hexasaccharides, 6,352 sequences or 92% of the total, have a GOLD score of 75 or less suggesting that these sequences minimally or do not bind antithrombin. Nearly 6.4% or 439 sequences have a GOLD score between 75 and 100. This score corresponds to μM to mM antithrombin affinity, most probably arising from non-specific, non-productive interactions.

The top 1% of the sequences or 28 sequences, each with a GOLD score greater than 100 were identified as 'hits' or high-affinity sequences. Of these 28 sequences, 23 were strikingly similar in possessing three sulfate groups, $6S_D$, $3S_F$, and $2S_F$. In addition, these 23 sequences had glucuronic acid residue in the 'E' position. This similarity is particularly significant considering that 513 sequences of the 6,859 hexasaccharides, i.e., ~7%, contain these structural features. In other words, the 'DEF' motif has been enriched from ~7% to ~82% through the affinity filter.
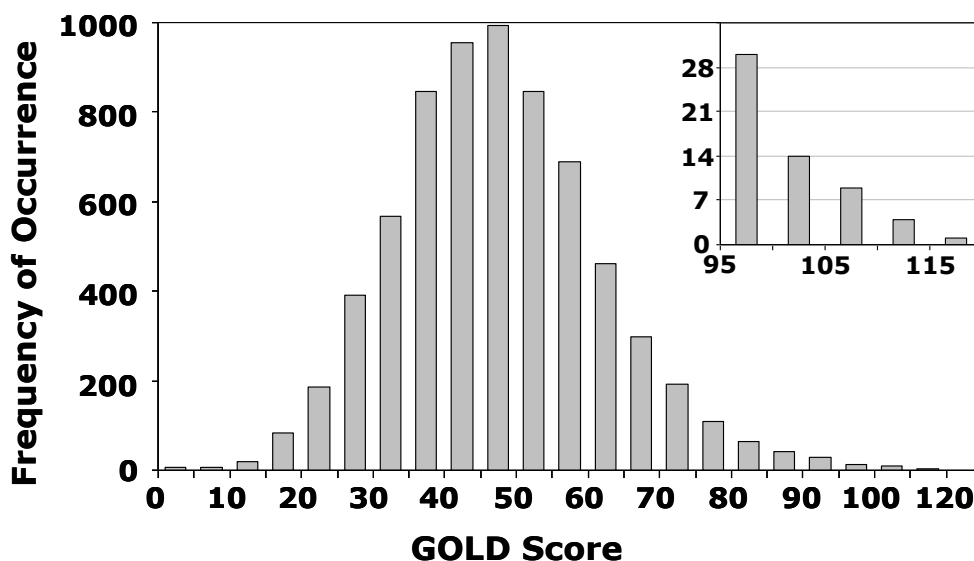
Figure 29. Histogram of number of HS hexasaccharide sequences for every 5 unit change in GOLD score. Modified GOLD score was calculated for all 6,859 hexasaccharides docked onto antithrombin following the first phase of combinatorial library screening. Inset shows an expanded portion of the histogram in the range 95–120 GOLD score.

The 28 'hits' identified in the affinity screen were then subjected to specificity analysis through a 'consistency test' with 100,000 iterations over triplicate docking runs. This step involved greater evolution of docking geometries, whereby significantly more conformational space around the binding site is searched multiple times. Of the 28 high-affinity sequences, 10 docked to consistently give a single geometry with an RMSD of less than 2.5 Å among its final solutions (Figure 30). The other 18 sequences did not dock consistently 100% of the time. The self-consistent binding geometry for each of the 10
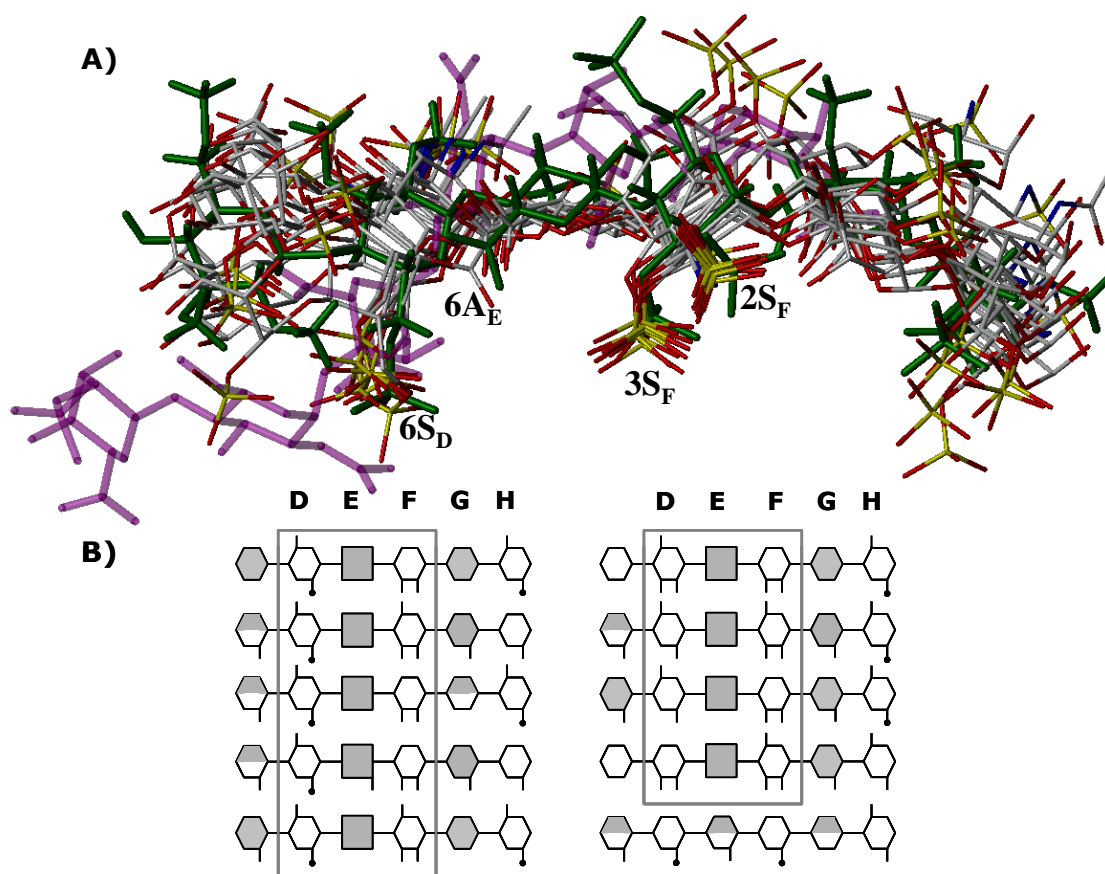
Figure 30. A) An overlay of final 10 hexasaccharide sequences obtained after second phase of combinatorial library screening. Structure in green is the H5$_{CRYS}$, while those in atom-type color are 9 sequences with nearly identical binding orientation and geometry. Sequence in purple color was found to bind antithrombin reproducibly with high specificity and affinity but dramatically different orientation. Labels 2S$_F$, 3S$_F$, 6A$_E$ and 6S$_D$ represent sulfate or carboxylate groups at the 2- and 3-position of residue F, 6-position of residue E and 6-positon of residue D. B) Symbolic representation of the high-affinity, high-specificity hexasaccharide structures shown above. The hexasaccharide library sequence runs {UAp(1→4)GlcNp(1→4)}$_3$, where UA is either IdoAp (shaded hexagon) or GlcAp (shaded square). Sulfated substitution at 2-, 3- or 6-positions of either UAp or GlcNp is indicated with a line (—), while acetate substitution at the 2-positon of GlcNp is indicated with a line-dot (—•). Iduronic acid residues in ^2S_O conformation are shown as fully shaded hexagons, while those in $^1C_4$ conformation are shown as half-filled hexagons.

sequences indicates a high specificity interaction. Of the 10 sequences, 9 bind antithrombin in manner identical to the natural pentasaccharide H5, while one hexasaccharide interacts with a significantly different geometry and orientation (Figure 30).

A closer look at these 9 sequences shows that 8 contain the →4) GlcNp6S (1→4) GlcAp (1→4) GlcNp2S3S (1→ structure that matches key structural features within DEF necessary for high-affinity interaction with antithrombin (Figure 7).[167,168,213] The last sequence in this category also has this structure, except for the absence of $6S_D$. Structural differences within DEF in these eight sequences include the presence of either an acetyl or a sulfate group at the 2 position of residue D and the presence or absence of a sulfate at 6-position of residue F. Our virtual screening shows that an IdoAp residue, either with or without a sulfate in 2-position, is consistently present at the G location in these 9 sequences (Figure 30). Alternatively, GlcAp residue is forbidden in this position. Further, the IdoAp residue overwhelmingly prefers ^2S_O conformation over $^1C_4$ (~89%) suggesting that the self-consistency filter is greatly sensitive to the local topology of the GAG helix. Finally, location H with GlcNp residue tolerates considerable variations, e.g., sulfation and acetylation in the 6- and 2-positions assuming that the high-affinity DEF sequence is present.

    The GOLD scores for the 9 hexasaccharide sequences ranged from 109.5 to 127.5 suggesting each of these sequences to be potent antithrombin ligands. Yet, despite having all the key structural features these are not as potent as sequence-specific pentasaccharide H5, which shows a GOLD score of 140. Comparison of the docking geometry of H5 with that of the hexasaccharide sequences reveals a lateral displacement of 0.3–0.5 Å in all

solutions irrespective of substitution pattern. This effect is most noted on G and H residues resulting in weaker interactions at the reducing end of the sequence. Although the small difference in binding geometry between pentasaccharide H5 and hexasaccharides remains unconfirmed, previous biochemical experiments have suggested a lateral movement as chain length increases.[212]

5.2.7. Identification of an unusual high-affinity, high-specificity sequence

Our combinatorial virtual screening approach results in one hexasaccharide sequence that binds antithrombin in an unusual, unexpected manner. This sequence, IdoAp2S (1→4) GlcNp2Ac (1→4) IdoAp2S (1→4) GlcNp2Ac (1→4) IdoAp (1→4) GlcNp2S6S (Fig. 8), does not have the critical DEF scaffold, is much less sulfated than H5, and has all three IdoAp residues in $^1C_4$ conformation. The sequence binds antithrombin reproducibly in triplicate docking experiments with a GOLD score of 109.0 indicating a high-affinity, high specificity interaction. The binding geometry is unique but distinct from H5 and other sequences identified above (Figure 30), specifically the sequence occupies part of both the pentasaccharide-binding site and the extended-heparin binding site in antithrombin. Significant binding energy originates from hydrogen bonding with amino acids Arg13, Arg46, Arg47, Lys125 and Arg129, in which the sulfate groups at the 2-position of the GlcNp and IdoAp residues make multi-valent interactions. At the same time, greater than 50% of its binding energy originates from van der Waals interactions. Thus, despite a lower than normal sulfation level, a fortuitous combination of

sequence and conformational features introduces high antithrombin affinity in this sequence.

It is difficult to predict whether this sequence will be physiologically active. The observation that less sulfated GAG sequence can demonstrate high antithrombin specificity *in silico* is exciting. Assuming that specific interactions of H5 are responsible for the allosteric conformational change in antithrombin, the dramatically different binding geometry suggests reduced likelihood of an agonist activity in this unusual sequence. At the same time, the high affinity of the sequence suggests an antagonistic activity may be expected.

5.3. Conclusion and Significance

Despite being the anticoagulant of choice for the past 8 decades, heparin-based therapy is still beset with several adverse effects, principal being an enhanced risk toward bleeding. A large body of structure-activity studies spanning more than two decades resulted in several potent pentasaccharides, of which fondaparinux was introduced in the clinic in late 2001.[26,146,211] Yet, these pentasaccharides, and also several low-molecular-weight heparins introduced in mid-1990s, are essentially 'heparin' and structurally represent a small advance.

Non-heparin anticoagulants with high specificity for antithrombin may afford a solution to current adverse effects of heparin therapy. Yet, designing non-heparin, organic activator(s) of antithrombin is challenging, especially due to the non-availability of computational techniques that address induced-fit mechanisms and specificity of binding

in highly-charged, solvent exposed systems.[171] Previous two simulations of heparin pentasaccharide binding to antithrombin were successful in identifying only few of the critical interactions.[205,214] A major reason for this is the difficult simulation of GAG interactions originating from the primary sequence complexity of GAG polymers, the inadequate parameterization of sulfate groups, and the shallow, exposed and highly flexible nature of GAG binding pockets.

Most docking approaches to date focus primarily on the affinity of interaction, and minimally on the specificity of interaction.[215] Predicting affinity of protein–ligand complexes accurately is in its infancy, especially due to poorly defined contribution of water molecules in the process. Further, affinity alone is not a powerful tool in predicting potent GAG sequences because of poor sensitivity to structural change, as current study (see Fig. 6) and work elsewhere demonstrate.[170] Instead of affinity, our genetic algorithm-based approach places emphasis on specificity of interaction. This approach affords screening of a large amount of conformational space that ensures elimination of a number of false positives. Thus, specificity is perhaps a more critical filter that offers significantly high reliability.

Our work represents the first approach of combinatorial library screening for heparin/HS GAGs. The phenomenal structural diversity of these GAGs is a challenge to medicinal chemists and represents a frontier extremely difficult to traverse through traditional structure-activity studies.[26] This work demonstrates that library screening is feasible for heparin/HS oligosaccharides, especially if a high-resolution crystal structure of the protein is available. Further, the approach is geared toward predicting high specificity

sequences. Thus, it is expected to be especially useful for heparin binding proteins, e.g., heparin co-factor II, protein C inhibitor and growth factors, for which specificity features remain poorly defined. In addition, for many other proteins, including glycoprotein D of herpes simplex virus–1 and protease nexin 1, information regarding heparin-binding site may also become feasible.

Our dual filter strategy – affinity and specificity – utilized fairly stringent criteria for selection. While the affinity filter selected the upper 20% of sequences, the specificity filter was set to select only those sequences that satisfy self-consistency 100% of the time. In combination, a high overall enrichment was realized (~660-fold). If one considers only those hexasaccharide sequences that possess IdoAp (1→4) GlcNp (1→4) GlcAp (1→4) GlcNp (1→4) IdoAp (1→4) GlcNp structure, which is the base sequence present in the final 9 hexasaccharides, even then the enrichment remains substantial (~77-fold). It was possible to use high filtering stringency because antithrombin–heparin interaction is biochemically a well-studied system. However, the stringent criteria may eliminate potentially useful information, especially for not as well-understood systems.

Our combinatorial screening approach relies on a fundamental hypothesis. The combinatorial library was generated based on the 'average backbone' hypothesis in which uniformity of inter-glycosidic torsion angles, irrespective of sequence and intra-ring conformational variability, is assumed.[109,110,113,114] For heparin/HS, this hypothesis appears to hold true for all known cases, although only a fraction of possible homogeneous sequences have been studied in solution. For GAGs other than HS, such as dermatan

sulfate and chondroitin sulfate that involve mixed inter-glycosidic linkages, the approach needs rigorous testing.

Although this approach is designed to identify individual GAG sequences with high specificity, it facilitates the extraction of a 'pharmacophore', key interactions that drive binding specificity. A map of RMSD of individual atoms, backbone as well as functional groups, can be readily created from the combinatorial library screening experiments to identify those locations with minimal variation, e.g., $3S_F$, $2S_F$, and $6S_D$ in DEF, which define the pharmacophore. In contrast, domains with higher RMSD, and therefore possessing significant movement, define locations in which structural modification can be introduced to design new ligands. Thus, this approach is also likely to be useful for designing therapeutically useful molecules.

In conclusion, the work describes identification of high-affinity high specificity heparin/HS sequences that bind antithrombin utilizing a combinatorial virtual library screening approach. The approach relies on a dual filter strategy involving affinity and specificity filters and is based on average heparin/HS backbone hypothesis. The approach uses genetic algorithm based docking and scoring protocol. The dual filter strategy, if found useful for other heparin/HS systems, is likely to be extremely useful in the design of pharmaceutically useful agents.

5.4. Methods

5.4.1. Software/Hardware

SYBYL 6.9.2 (Tripos Associates, St. Louis, MO) was used for molecular visualization, minimization and for adding hydrogens to protein structures from the Protein DataBank. All modeling was performed on MIPS R16K or R14K IRIX 6.5-based SGI Tezro and Fuel graphical workstations. GOLD[TM] (version 2.2, Cambridge Crystallographic Data Center) was used for docking experiments.[207] Combinatorial GAG structures were built in an automated manner using in-house SPL (SYBYL Programming Language) scripts.

5.4.2. Energy Minimizations

Energy minimization of modeled structures was performed to optimize the geometric conformation of GAG and AT. Except where stated, energy minimization was performed using the Tripos forcefield with Gasteiger-Hückel charges, a fixed dielectric constant of 80 and a non-bonded cutoff radius of 8 Å. Minimization was carried out for a maximum of 5,000 iterations subject to a termination gradient of 0.05 kcal/(mol·Å).

5.4.3. Protein Co-ordinates

The coordinates for the activated form of AT were extracted from the crystal structure of the ternary AT-pentasaccharide-thrombin complex (PDB entry 1TB6).[119] Hydrogen atoms were added in SYBYL 6.9.2 and the structure minimized with fixed heavy-atom co-ordinates using the Tripos forcefield for 1,000 iterations subject to a

termination gradient of 0.05 kcal/(mol·Å). Single-point mutants of AT, including K114Q, K125M and R129H, were generated using the residue mutation protocol in SYBYL. The side-chains of the mutated residues were optimized through energy minimization in which the all other side chains were held rigid.

5.4.4. Co-ordinates for synthetic pentasaccharide H5$_{CRYS}$

The 1TB6 antithrombin–heparin–thrombin ternary crystal structure[119] has a heparin chain containing a synthetic pentasaccharide sequence, H5$_{CRYS}$, which is a variant of the natural pentasaccharide DEFGH (Figure 24). This synthetic pentasaccharide H5$_{CRYS}$ has six *O*-sulfate groups in residues D (position 6), F (positions 2,3 and 6) and H (positions 2 and 6), has glucose residues, instead of glucosamines, and has all available –OH groups in residues D (positions 2 and 3), E (positions 2 and 3), G (positions 2 and 3) and H (positions 1 and 3) protected in the form of –OCH$_3$ groups. The atom type of sulfur and oxygen atoms in –SO$_3$ groups was modified to S.o2 and O.co2, respectively, and the bond type between these atoms was modified to aromatic bond. Hydrogen atoms, absent in the PDB structure, were added in SYBYL and the resultant structure minimized to optimize geometry of hydrogen atoms only (no change in non-H atoms).

5.4.5. Natural Pentasaccharide DEFGH co-ordinates

The sequence-specific pentasaccharide H5, or sequence DEFGH (Figure 7A), was modeled by introducing the necessary changes to the 1TB6 pentasaccharide. The methylated anomeric reducing end was retained to simulate the environment of the +1 residue. Several research groups have reported ϕ_H/ψ_H inter-glycosidic torsion angles for

HL-GAGs in the free and bound conformations.[109-111,113,114,118] These torsion angles fall within a relatively narrow range and suggest that they remain fairly constant irrespective of a change in substitution pattern around the inter-glycosidic bond. Thus, we utilized an average value of bond torsion for each inter-glycosidic linkage, which was the mean of the two extremes reported in literature. The natural sequence H5 was minimized at the average ϕ_H/ψ_H values, subject to a restraining force constant of 0.01 kcal·mol^{-1}·deg^{-2}. The final ϕ_H/ψ_H values, following minimization, deviated by not more than $\pm7°$ from the initial values. The minimized structure retained the initial ^2S_O IdoAp conformation and was used for docking studies.

5.4.6. Co-ordinates for Variant Pentasaccharides

The 'average-backbone' natural pentasaccharide DEFGH structure was used as a template for the generation of H5 variants. These variants include truncated forms, tetrasaccharides H5-H and H5-D, trisaccharides H5-GH and H5-DE, and functional group variants H5+3S$_H$, H5+3S$_E$, H5+3S$_G$, H5-6S$_D$, H5-3S$_F$ and H5-2S$_H$ (Figure 26). For each of these structural variants, following appropriate structural modifications in the natural pentasaccharide sequence, minimization was performed with constraints that retain the average ϕ_H/ψ_H values. In addition to these structural variants of the natural sequence, a DEFGH sequence was also prepared in which the iduronic acid residue G was in the $^1C_4$ conformation, labeled as H5/G*, rather than the ^2S_O form. The $^1C_4$ IdoAp ring co-ordinates were taken from the crystal structure of FGF2-GAG (PDB entry 1BFC)[53] and the structure minimized as described above.

5.4.7. Co-ordinates for HL-GAG Virtual Library

The co-ordinates for the HL-GAG sequences of the virtual HS hexasaccharide library were generated with a series of SPL scripts and a set of nineteen HS disaccharide building blocks (Figure 31, below). Although the number of possible HL-GAG UAp(1→4)GlcNp disaccharides is 48 (Figure 4), only 23 have been experimentally observed. Based on sequence-specific natural pentasaccharide sequence, we restricted our library to include GlcAp sequences that have GlcNp3S and IdoAp sequences that do not contain GlcNp3S. Because IdoAp residues in heparin can exist either in the ^2S_O or $^1C_4$ conformations, each IdoAp residue was modeled explicitly in these two different states. Thus, our virtual library consisted of 16 IdoAp- and 3 GlcAp-containing disaccharides generating a total of 19 building blocks.

GlcAp- and ^2S_O-IdoAp-containing disaccharides were generated using the EF and GH residues from the 1TB6 co-crystal structure as template, while the template for the $^1C_4$-IdoAp disaccharides was obtained from the 1BFC structure.[53] Appropriate side-chain modifications were made to generate the 19 building blocks. Each disaccharide was minimized at the average ϕ_H/ψ_H value subject to a restraining force constant of 0.01 kcal·mol^{-1}·deg^{-2}. The 19 disaccharides were then used to build a combinatorial HS hexasaccharide library using an SPL script, following which each sequence was minimized as described above in an automated manner. Thus, the HS combinatorial library contained $19 \times 19 \times 19 = 6,859$ hexasaccharide sequences.

5.4.8. Docking of HL-GAG Sequences

Docking of saccharide ligands onto the activated form of antithrombin was performed with GOLD[TM] v.2.2. The binding site in antithrombin was comprised of all atoms within 16 Å from the C^ζ atom of Phe121 in the D helix. This dimension of the binding site covers all important known heparin-binding residues including Lys11, Arg13, Arg46, Arg47, Trp49, Lys114, Phe121, Lys125, Arg129, and Arg132.[25,118,167] GOLD[TM] is a "soft docking" method that implicitly handles local protein flexibility by allowing a small degree of interpenetration, or van der Waals overlap, of ligand and protein atoms. GOLD also optimizes the positions of hydrogen-bond donating atoms on Ser, Thr, Tyr and, most importantly, Lys residues as the part of the docking process. Whereas all saccharide bonds were constrained for the rigid body docking experiment, only the inter-glycosidic bonds were constrained when docking structures with the average torsion angles.

For the native H5 sequence, and its truncated and variant forms (Figure 26), docking was performed using a genetic algorithmic search with 100,000 iterations for each of 10 runs.[207] In this search, GOLD[TM] starts with a population of 100 arbitrarily docked ligand orientations, evaluates them using a scoring function (the GA "fitness" function) and improves their average "fitness" by an iterative optimization procedure that is biased towards high scores. As the initial population is selected at random, several such GA runs are required to more reliably predict correct bound conformations. In this study 10 GA runs were performed with the GOLD score as the "fitness" function. Collectively, these 10 GA runs will be referred to as one docking experiment. In addition, to enhance speed, the GA was set to pre-terminate if the top two ranked solutions were within 2.5 Å RMSD.

Docking experiments were performed in triplicate to ensure reproducibility and to reduce false positives. The top two solutions of each docking experiment were considered for further analysis. Thus, a typical triplicate docking experiment would yield a minimum of 6 solutions.

In contrast, when docking the HS combinatorial library made from 19 disaccharides (Figure 31, below), a two-step docking protocol was utilized. The first step consisted of screening all possible sequences using 10,000 GA iterations and GOLD score evaluation of only the top-ranked solution. This step identified the most promising sequences (~ top 1%) that have a relative high GOLD score. The second step consisted of docking these most interesting sequences according to the protocol described above for the natural pentasaccharide DEFGH and its variants.

Docking was driven by the GOLD scoring function. Although this scoring function correlates with the observed free energy of binding, a modified form of the scoring function has been found to be more reliable.[50] This modified GOLD score, which utilizes hydrogen-bonding and van der Waals interactions (Equation III), was used to rank the final docked solutions.
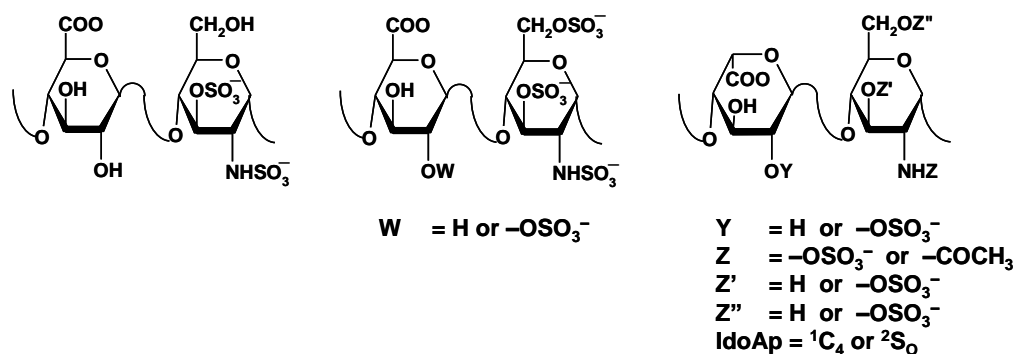
Figure 31. Disaccharide sequences used to build virtual library of 6,859 hexasaccharides.

.

$$GOLD \cdot Score = HB_{EXT} + 1.375 \times VDW_{EXT} \qquad\qquad [III]$$

where HB_{EXT} and VDW_{EXT} are the "external" (nonbonded interactions taking place where HB_{EXT} and VDW_{EXT} are the "external" (nonbonded interactions taking place between the ligand and receptor) hydrogen bonding and van der Waals terms, respectively

THE INTERACTION OF HEPARIN / HEPARAN SULFATE WITH

HEPARIN CO-FACTOR II — A HYPOTHESIS

6.1. Introduction

Heparin cofactor II (HCII) and antithrombin (AT) are *ser*ine *p*roteinase *in*hibitors (serpins) present in human plasma at significant levels (1.2 and 2.3 μM, respectively).[216] AT and HCII bind heparin (H) and heparan sulfate (HS), two important members of the glycosaminoglycan (GAG) superfamily. Whereas the interaction of AT with H/HS has been the basis for several pharmaceutical agents, e.g., full-length heparin, low molecular weight heparins and fondaparinux, the interaction of HCII with H/HS has not been exploited for drug design to date.

AT binds to a five-residue sequence in H/HS – the so-called heparin pentasaccharide sequence, H5 (Figure 7) – with high affinity in the region formed by helices A and D.[118] This high affinity interaction is known to be highly specific. Of the 11 positively charged amino acid residues in this heparin-binding site (HBS) in AT, Lys114, Lys125 and Arg129 have been found to be critical for recognition, affinity and specificity of the H5 sequence.[167,190-192] Recently solved crystal structures[34] of HCII in its native and thrombin-complexed states show a striking similarity with AT in the organization and

100

orientation of several positively charged amino acid residues in the putative HBS, e.g., Arg103, Lys173, Lys185, Arg189, Arg192, and Arg193. Yet, the heparin pentasaccharide sequence does not bind HCII with high affinity and specificity.[217] In fact, the interaction of heparin cofactor II with heparin and heparan sulfate is believed to be non-specific.[32] In addition to the structural similarities between the two serpins, mechanistic similarities also exist. Both serpins undergo a two-step, induced-fit, allosteric activation mechanism to inhibit the target enzyme.[27,32,168] In this mechanism, an initial loose recognition complex is first formed, which undergoes a rapid conformational change in the second step to form a tight, high-affinity heparin-serpin complex. This conformational change is the basis for the ~300-fold activation of AT in its inhibition of factor Xa[27,168] and the ~1000-fold activation of HCII in its inhibition of thrombin.[32] Yet, whereas the pentasaccharide sequence (Figure 7) in H/HS specifically performs this activation for antithrombin, the sequence(s) that activate HCII remain(s) unknown.

The identification of H/HS sequences that bind HCII with high affinity and high specificity could have major advantages. In contrast to AT, which inhibits several proteinases of the coagulation cascade, e.g., thrombin, factor Xa and factor IXa,[11,218] HCII inhibits thrombin only.[12] This intrinsic specificity may be a unique advantage because HCII deficiency does not appear to enhance risk for thrombosis,[21] while studies suggest that the serpin may play a role in preventing arterial thrombosis.[20,21,219,220] Further, HCII is able to inhibit clot-bound thrombin, in striking contrast to antithrombin.[221] Thus, an HCII-based anticoagulant may serve as a much needed potent, yet safe, regulator of clotting.

To understand H/HS – HCII interaction, we utilized our dual-filter genetic algorithm-based combinatorial virtual screening approach, which we developed to deduce H/HS sequences that recognize AT.[5] Using this dual-filter strategy, we have identified five H/HS sequences from a combinatorial library of 46,656 H/HS hexasaccharides that are predicted to bind HCII with 'high affinity and high specificity'. Contrary to the common thinking, our results suggest that discrete sequences in H/HS may exist for specific interaction with HCII. Our computational work highlights a critical need for detailed solution molecular interaction studies using chemically synthesized, homogeneous H/HS sequences, which may pave the way for new, specific HCII-based anticoagulants.

6.2. Results and Discussion

6.2.1. Structure of the Activated Form of Heparin Cofactor II

Two experimentally determined structures of HCII are available, native and S195A thrombin-complexed.[34] The overall structure of native HCII is similar to that of native AT. Superposition of the structure of native HCII on that of native AT (PDB file: 2ANT)[188] gives a RMSD of 1.8 Å for 352 equivalent C^{α} atoms (not shown). Likewise, superposition of C^{α} atoms of the residues that define the HBS in AT, i.e., Arg46, Arg47, Lys114, Lys125, Arg129, Arg132 and Lys133, with corresponding residues in HCII results in a RMSD of 1.5 Å indicating a high degree of similarity between the two native serpins.

The structure of GAG-activated HCII is not available as yet. However, the structure of the serpin in complex with S195A thrombin displays extensive similarities with that of

the heparin pentasaccharide-activated AT (Figure 32). The S195A thrombin-complexed HCII shows an expelled reactive center loop (RCL), as found in pentasaccharide-activated AT.[118,119] The reason for the expulsion of the RCL appears to be the extensive exosite interactions that the RCL makes with S195A thrombin. Likewise, exosite interactions also stabilize the heparin-induced, conformationally activated AT, as borne out in experiments with factor Xa and factor IXa.[222,223] Thus, the overall structure of HCII in the S195A thrombin-complexed state is similar to that of the pentasaccharide-activated AT.

Further evidence that these forms are nearly identical comes from the superposition of the corresponding C^α atoms of the activated serpins. Figure 32 shows the superimposed activated forms of the two serpins. The RMSD in C^α atoms of the core amino acid residues was found to be 2.4 Å suggesting significant similarity in the orientation of most structural domains including β-sheets and helices. More importantly, the RMSD for corresponding basic residues in helices A and D was found to be 1.5 Å indicating a high degree of similarity between the two activated forms in this region. These structural similarities indicate that the S195A thrombin-complexed HCII is likely to be the H/HS-activated form of the serpin.

Despite the structural similarity, fine differences exist between the two activated serpins in the relative orientation of helices A and D, and in the extension of RCL. Whereas A helices superpose nearly completely, the D helices display a significant ~30$^\text{O}$ angle between the two serpins (Figure 32). Likewise, the RCL in the activated forms show a significantly greater extension of the loop in the S195A thrombin-complexed HCII than
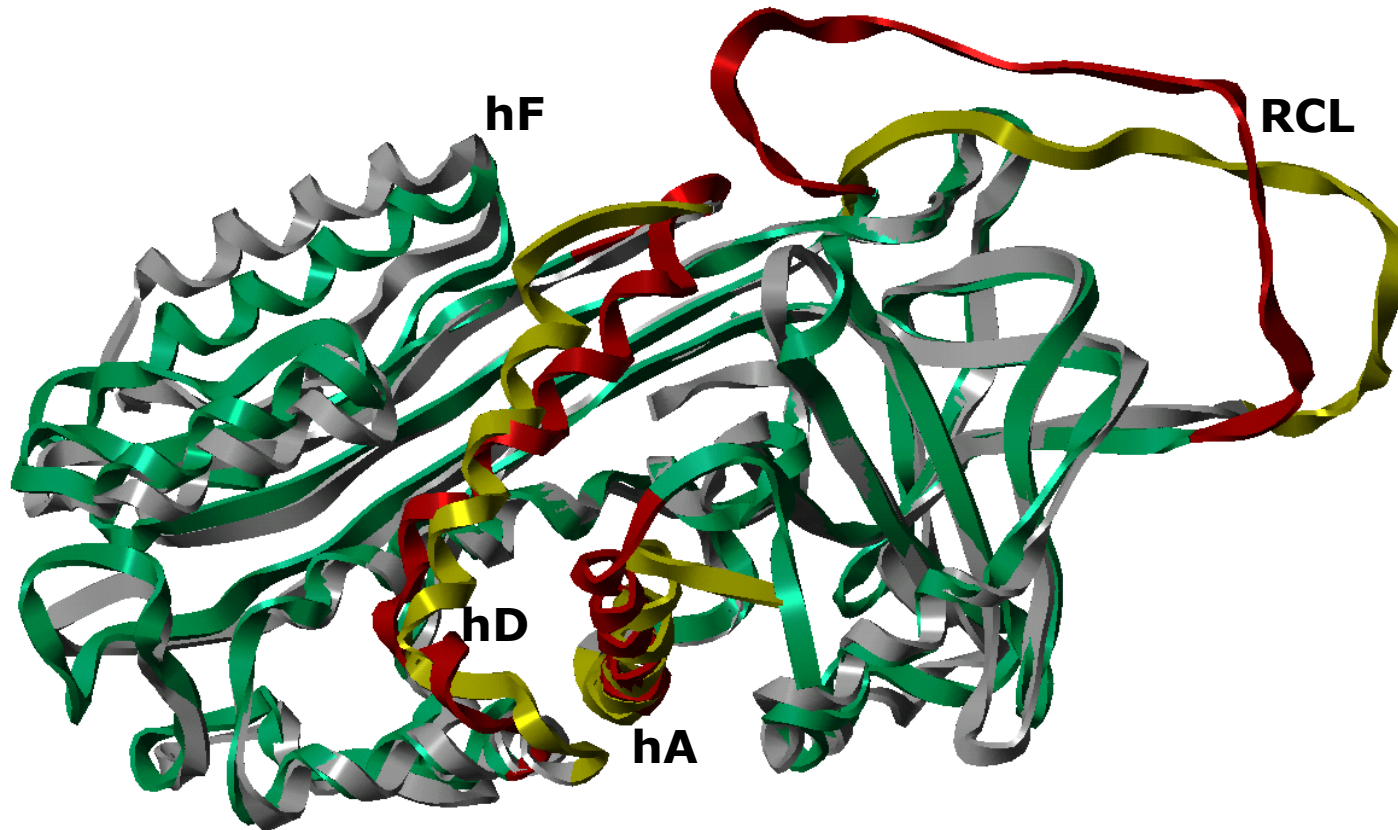
Figure 32. Comparison of the structure of S195A thrombin-complexed heparin cofactor II with heparin pentasaccharide activated antithrombin. Core polypeptide sequences, devoid of residues of the *N*-terminus and the RCL, were aligned using Sybyl 7.2 homology fit algorithm. Note the small rotational difference in helix D axis and at the *N*-terminal end of helix A between the two proteins, while the expulsion of RCL in HCII is much greater than that in AT. Antithrombin ribbon is shown in green and red (hD, hA and RCL), while HCII is shown in gray and yellow (hD, hA and RCL).

in the pentasaccharide-activated AT. In addition, helix D of HCII contains an additional electropositive residue, Arg184, which has no counterpart in AT. Despite these differences, the S195A thrombin complexed HCII structure is very similar to the pentasaccharide-activated AT structure and is a good model for investigation of H/HS interactions, especially considering that the structure of GAG-activated HCII is unknown.

6.2.2. Rapid filtering of sub-optimal HS sequences from a library of 46,656 sequences

Recently, we designed a genetic algorithm-based virtual screening approach that utilized a dual-filter process to identify hexasaccharide sequences in heparin that recognize AT with high specificity.[5] This sequential dual-filter algorithm utilizes GOLD score, a measure of 'affinity', as the first filter, followed by convergence of binding geometries, a measure of 'specificity', as the second (Figure 33). The dual-filter algorithm predicted ten 'high-affinity and high-specificity' AT-binding hexasaccharide sequences from a combinatorial library of ~7,000 sequences. Nine of the 10 sequences contained structural features matching the high-affinity heparin pentasaccharide sequence shown in Fig. 1A. Additionally, the computationally predicted binding geometry matched the crystal structure geometry to within 2.5 Å.[5]

In the present study with HCII, we used a comprehensive library of H/HS hexasaccharide sequences built from all of the 23 disaccharide building blocks reported to date (Figure 4). To address the conformational variability possible in IdoAp residues, two major ground state conformers $^{1}C_{4}$ and $^{2}S_{O}$, were explicitly modeled, increasing the
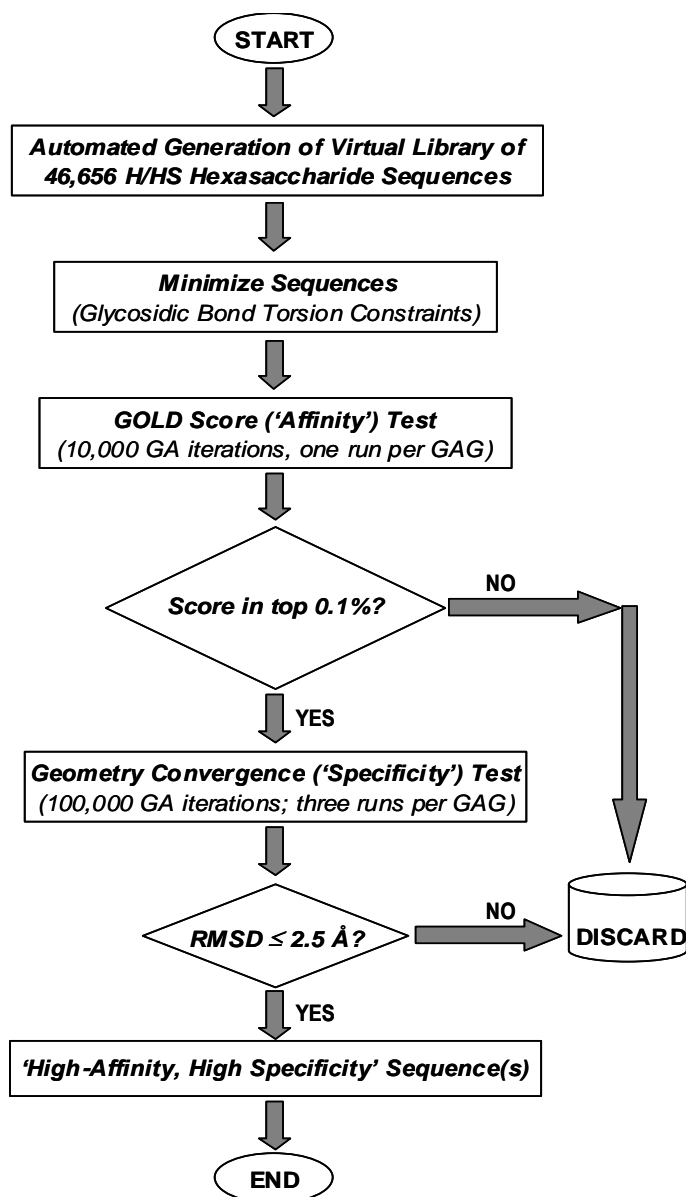
Figure 33. Dual-filter algorithm used to screen a combinatorial library of 46,656 H/HS

hexasaccharide sequences. The combinatorial library was built from 36 naturally occurring

disaccharide building blocks and screened using a dual filter protocol.
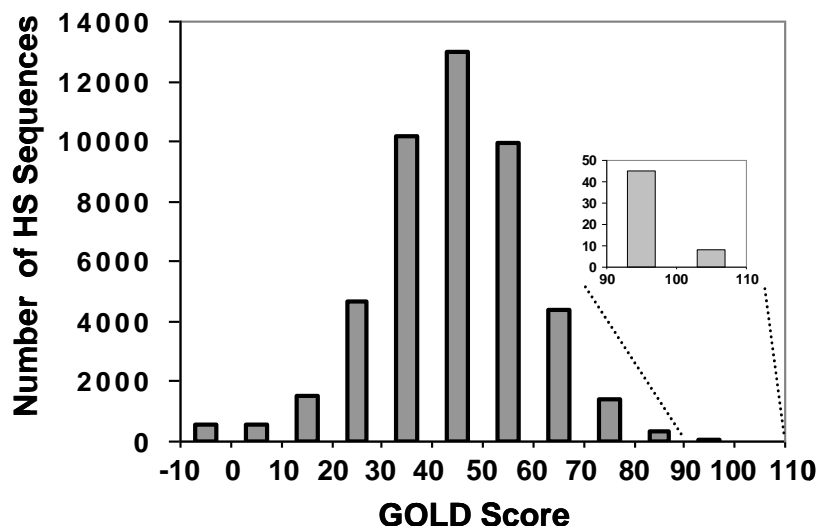
Figure 34. Histogram of number of H/HS hexasaccharide sequences for every 10 unit change in GOLD score. shows an expanded portion of the histogram in the range 90–110 GOLD score units.

number of building blocks to 36 to give a library of $36^3 = 46,656$ hexasaccharide sequences. The H/HS binding site in activated HCII was defined to include the domain formed by helices A and D[34] considering the similarity with the H – AT system. Figure 33 displays the histogram of GOLD scores following the first step of the dual filter protocol. The profile is Gaussian and shows that a majority of H/HS sequences (83.5%) bind HCII with an average GOLD score in the region of 30 – 80 units. Nearly 15.7 % of sequences bind poorly (GOLD score below 30 units), while 0.8 % hexasaccharides recognize the serpin with high GOLD scores between 80 – 106 units. This included 2 sequences with GOLD score higher than 100 and 45 sequences with score between 90 and 100 units. In comparison to the H – AT system,[5] these overall GOLD scores are approximately 20–30

units lower suggesting that the affinity of H/HS hexasaccharides for HCII may be slightly lower than that for the high-affinity pentasaccharide binding to AT.

6.2.3. Structural features of the 'high-affinity' H/HS hexasaccharides

Structural analysis of the sequences identified through the 'GOLD score' filter reveals interesting insight into recognition of HCII. Of the 47 hexasaccharides, only 4 carry the maximal possible sulfation load of 9 groups, while the library contains a total of 1728 sequences with the maximum load. This suggests that the majority of the high-affinity sequences that bind HCII are not highly sulfated. Alternatively, it suggests that the GOLD fitness function does not arbitrarily select for higher sulfation level, although the binding site in HCII is electropositive. Eight topologies corresponding to the common heparin hexasaccharide sequence, [IdoAp2S-GlcNp2S6S]$_3$, which carry maximal sulfation load, have GOLD scores in the range of –2.7 to 66.4 units, suggesting that these sequences recognize HCII with poor affinity. These results are consistent with solution experiments with heterogeneous, polydisperse heparin sample that show poor HCII affinity in the range of 45 – 140 μM.[32,224,225]

The ratio of IdoAp- and GlcAp-containing disaccharides in our combinatorial library was 2.6:1, while it was found to be 1.2:1 in the 47 hexasaccharides. This implies a significant enrichment of GlcAp-residues. Additionally, the top 0.1 % identified 'hits' (47 topologies) do not contain any sequence related to the high-affinity pentasaccharide sequence (IdoAp2S {^{2}S$_O$ / ^{1}C$_4$}-GlcNp2S6S-GlcAp-GlcNp2S3S6S-IdoAp2S {^{2}S$_O$ / ^{1}C$_4$}-GlcNp2S6S) that binds AT. This result is also consistent with Maimone and Tollefsen,

who have shown that heparin molecules with or without the high-affinity antithrombin-binding pentasaccharide sequence equally activate HCII.[217]

6.2.4. Finding Needle(s) in the Haystack: Only five H/HS hexasaccharides are predicted to recognize that activated form of HCII with 'high-specificity'

The geometry convergence filter used in our dual-filter algorithm is a robust strategy to identify sequences that possess exceptional complementarity to the receptor. This filter utilizes 3 experiments of 10 GA runs each, in which each GA run is allowed to evolve over 100,000 iterations, resulting in 6 final binding geometries. Binding is deemed to be 'specific' if the RMSD among these 6 geometries is ≤ 2.5 Å. Of the 47 sequences that were subjected to this stringent criterion, only 5 sequences were found to recognize activated HCII with 'high specificity' (Figure 35).

Several points about the structure of these five hexasaccharide sequences are striking. None of the GlcNp residues have an acetyl group at the 2-position. This is striking because natural heparan sulfate consists of nearly 50–60 % GlcNp2Ac residues.[226,227] The total number of sulfate groups in these sequences ranges between 6 (sequence **I**) to 8 (sequence **V**). This averages to about 2.0 to 2.6 sulfate groups per disaccharide sequence. In contrast, the degree of sulfation of human liver HS and porcine liver HS is 1.2 and 1.0, respectively, while that of porcine intestinal heparin is 2.6.[226-228] Thus, these sequences are significantly more sulfated than natural HS, but either equal to or slightly less sulfated than natural heparin. Yet, these sequences are unlike natural H because the IdoAp composition (20%) is dramatically lower than that expected of natural heparin (>80%).[1]
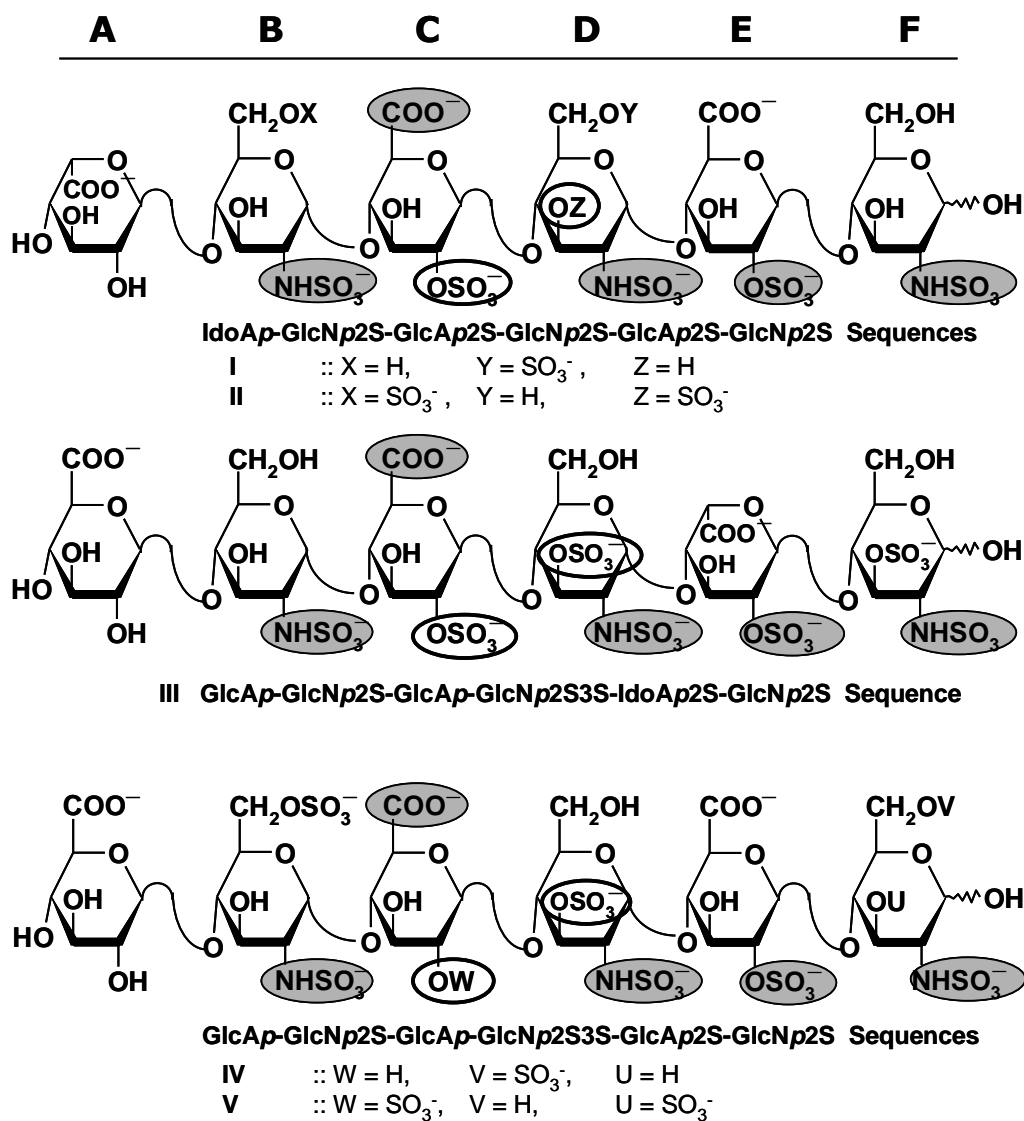
Figure 35. Structures of five H/HS hexasaccharides, sequences I through V, which are predicted to recognize HCII with 'high affinity and high specificity'. Each sequence contains more than one residue that is rarely found in naturally occurring HS.

An IdoA*p* residue is present at the non-reducing terminus in sequences **I** and **II**, and at an internal position – residue E – in sequence **III**. Each of these IdoA*p* residues has the ^2S_O conformation. At the remaining twelve locations for uronic acid residue, a GlcA*p* residue is present. In addition to these differences, variations in the presence or absence of sulfate group(s) are found in ring B, C, D, or ring F (Figure 35). Despite these differences, closer inspection reveals striking similarity. The fundamental structure that is common to all five sequences is UA*p*–GlcN*p*2S–GlcA*p*–GlcN*p*2S–UA*p*2S–GlcN*p*2S. Further, the 2- and 3-positions of rings C and D bear a sulfate group in 4 out of 5 sequences highlighting their importance in HCII recognition.

6.2.5. Molecular Interaction Profile of the Five Predicted 'High-Affinity, High-Specificity' Sequences

When modeled in complex with activated HCII, all five sequences bind in an essentially identical orientation with the non-reducing end recognizing helix A, while the reducing end is oriented toward the *C*-terminus of helix D (Figure 36). This suggests exquisite specificity in the recognition of activated HCII. In fact, a RMSD of 0.76 Å is found for backbone atoms of the central tetrasaccharide BCDE among the top scoring solutions for each of the five sequences, which increases to 1.6 Å, if one considers the backbone atoms of all six residues. The H/HS hexasaccharides orient at a ~60º angle relative to the axis of helix D. This alignment is significantly different from heparin pentasaccharide binding to AT, which orients almost parallel to the axis of helix
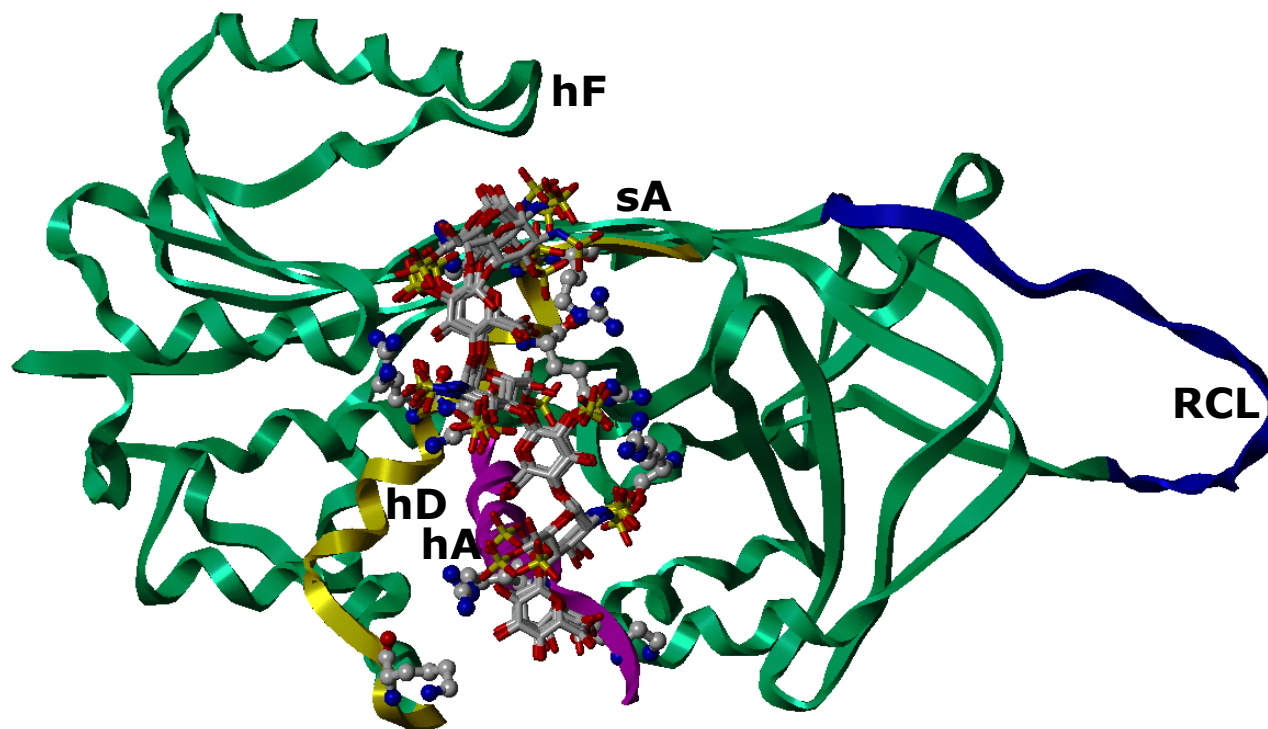
Figure 36. Combinatorial virtual screening predicted interaction of H/HS with HCII. Ribbon diagram of HCII, from S195A thrombin-complexed HCII, interacting with five H/HS sequences I through V of Figure 35. The top scoring solutions from the first docking run are shown. H/HS hexasaccharides (shown in capped sticks) bind in the domain formed by helix D (shown in yellow) and *N*-terminal residues of helix A (magenta). Prominent basic residues are shown as ball and stick. Note that the four central residues – BCDE – superimpose with almost no deviation. Reactive center loop (RCL) is shown in blue. hF = helix F; sA = β-sheet A.
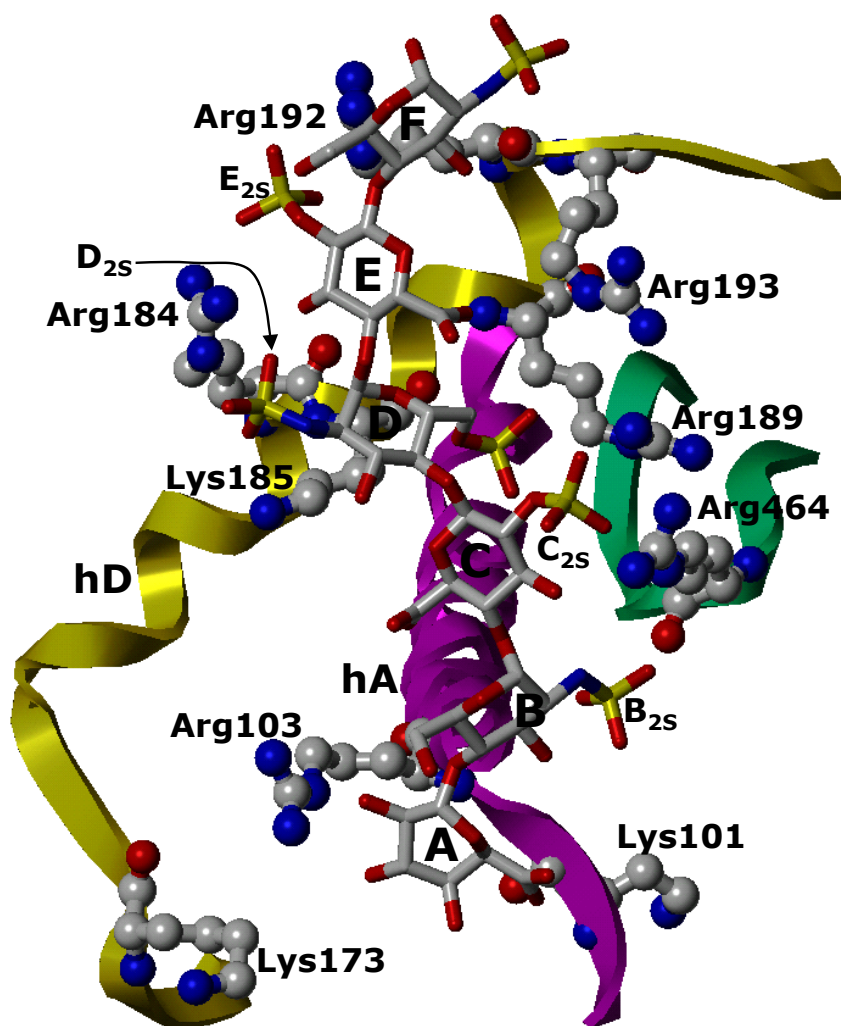
Figure 37. A close-up view of the HS hexasaccharide sequence I, with highest GOLD score, binding to activated HCII.

D.[118,119]Three residues, Arg184, Arg192, and Arg464, form strong hydrogen bonds with all five sequences (Figure 37), while Arg193 and Arg189 formadditional hydrogen bonds with two and one of five hexasaccharides, respectively (not shown). No hydrogen bonds are predicted to form with Lys101, Arg103, Lys173 and Lys185, other basic residues present in the binding site. Correlations can be found between HCII site-directed mutagenesis

results and our computational results. For example, affinity chromatography studies suggest that Arg184, Arg192 and Arg193 of HCII interact with heparin, while Arg103 does not.[229,230] Our computational study supports these observations. In contrast, studies with Lys173 mutant show that this residue may bind H/HS,[231] while none of our final H/HS hexasaccharides appear to interact with Lys173. It is possible that chains longer than hexasaccharides may be needed because Lys173 is present ~2.0 Å further from the non-reducing end of hexasaccharide sequences **I** to **V**.

Closer inspection of interaction at the atomic level reveals that the high scoring residues – Arg464, Arg184 and Arg192 – form multi-valent hydrogen bonds (Figure 37). Specifically, Arg464 forms 3 strong hydrogen bonds, two with the B_{2S} group and one with the C_{2S} group, Arg184 forms 2 hydrogen bonds with the D_{2S} and E_{2S} groups, while Arg192 forms 1 – 2 hydrogen bonds with E_{2S} group. Together these four sulfate groups on the GAG are responsible for over 85% of the calculated hydrogen bonding score (not shown). This analysis suggests that a core, conserved tetrasaccharide BCDE with the minimal sequence GlcN*p*2S—GlcA*p*2S—GlcN*p*2S—UA*p*2S appears to be critical for 'high-affinity, high-specificity' binding to activated HCII. Likewise, sulfations at 3- and 6-positions in residues F and D, respectively, introduce additional interactions with Arg193 and Arg189, respectively, and are therefore desirable.

6.3. Significance

The interaction of H/HS with HCII has been commonly assumed to be non-specific[32,217,225] primarily because natural GAG preparation could not be separated into

high and low affinity fractions. This also poses major difficulties in structure elucidation because crystallization of a co-complex using heterogeneous, polydisperse GAG is nearly impossible. Our work suggests that there exist at least five hexasaccharide sequences that are predicted to recognize activated HCII with 'high-affinity and high specificity'. All five 'needles in the haystack' contain at least two GAG monosaccharides of the GlcN*p*2S3S or GlcA*p*2S type that are rare in natural preparations of H/HS. The fact that each sequence has at least two such rare monosaccharides (sequences **II** and **III** have three, while sequence **V** has four) suggest a rapidly diminishing probability of finding these structures naturally. H/HS sequences **I** – **V** are expected to be accessible through chemical synthesis,[157] which may rejuvenate interest in understanding structure – function relationships in this system as well as advance the concept of designing potent HCII agonists as specific inhibitors of thrombin. Finally, considering that the structure of H/HS – heparin co-factor II complex is not available, our combinatorial virtual library screening results have far reaching implications.

6.4. Computational Methods

6.4.1. Protein Co-ordinates

SYBYL 7.2 (Tripos Associates, St. Louis, MO), working on MIPS R16K or R14K IRIX 6.5-based SGI Tezro and Fuel graphical workstations, was used to model protein structures. The coordinates for the activated form of HCII were extracted from the crystal structure of the S195A thrombin-HCII Michaelis complex (protein data bank (PDB) entry 1JMO), while the coordinates for native HCII were obtained from PDB entry 1JMJ.[34]

Following addition of hydrogen atoms, the protein structures were minimized with fixed heavy-atom co-ordinates using the Tripos Force Field for 1,000 iterations subject to a termination gradient of 0.05 kcal/(mol-Å). For comparison of protein structures (HCII complexed with S195A thrombin (1JMO) and activated form of AT (chain I of 1AZX)[118]), the *N*-terminal end residues 61 through 100 in HCII and 2 through 25 in AT, as well as reactive center loop residues 428 to 446 in HCII and 378 to 399 in AT, were deleted. The main bodies of the two serpins (the core residues) were then aligned using homologous residues in an automated manner and RMSD between alpha carbons of corresponding residues calculated.

6.4.2. Co-ordinates for H/HS Virtual Library

The co-ordinates for H/HS hexasaccharide sequences present in the combinatorial virtual library were generated using a series of in-house SYBYL programming language scripts. See section 5.4.7 for further details

6.4.3. Docking of the H/HS Virtual Library onto HCII

Docking of saccharide ligands onto the native and activated form of HCII was performed with GOLD v.2.2 (Cambridge Crystallographic Data Center, Cambridge, UK).[207] The substituents on the pyranose rings of H/HS sequences were allowed full conformational mobility in docking experiments. The binding site of H/HS sequences in native and activated HCII comprised of all atoms within 18 Å from the C^γ atom of Lys185 in the D helix, which covers Lys101, Arg103, Lys173, Arg184, Lys185, Arg189, Arg192, Arg193, Lys220, and Arg464 residues that are likely to form the putative HBS in HCII

(helices A and D). Unless otherwise specified, default parameters were employed during the GOLD docking run.[5]

The docking protocol was essentially similar to that used for antithrombin (section 5.4.8). Evaluation of the H/HS combinatorial hexasaccharide library was performed using the dual filter docking protocol (Figure 33) previously used in our study of the AT – heparin pentasaccharide system.[5] The first stage (the 'affinity' test) involved docking of 46,656 H/HS sequences to HCII using 10,000 iterations per GA run. The high affinity H/HS sequences so identified (47 sequences or ~0.1% of the total) were then selected for the geometry convergence ('specificity') test, which consisted of slower and more rigorous experimentation with 100,000 GA iterations performed in triplicate. The top two solutions of each docking experiment, i.e., 6 solutions, were retrieved and analyzed for RMSD among the backbone heavy atoms (pyranose ring atoms and interglycodic oxygens). An RMSD of less than 2.5 Å between the 6 solutions suggested a high degree of convergence to a 'unique' binding geometry.

DESIGN, SYNTHESIS AND EVALUATION OF POTENTIAL NEXT GENERATION ANTITHROMBIN ACTIVATORS

7.1. Introduction

Previously synthesized non-saccharide antithrombin (AT) activators in our laboratory namely sulfated flavonoids and sulfated tetrahydroisoquinoline derivatives, were found to bind the extended heparin-binding site (EHBS) or elsewhere although they were designed to target the pentasaccharide-binding site (PBS). The X-ray crystal structures of antithrombin alone[188,232] and in complex with heparins[118,119] show that the PBS is exposed to solvent implying that the binding domain should be freely available. Yet, biochemical studies in solution show that the *N*-terminus of the polypeptide overlays on the PBS.[233] This implies that the three key residues, Lys114, Lys125 and Arg129, are not readily accessible to ligands in solution. Thus, it is likely that sub-optimal activators that cannot engage all three key residues find it difficult to form a productive PBS-based antithrombin – ligand initial recognition complex,[192] which can initiate the induced-fit conformational change in the serpin. Rather, the sub-optimal activators are ensnared by an adjacent electropositive domain, the EHBS, resulting in lower activation. Thus, a plausible non- saccharide design strategy would be to devise larger anionic molecules, which upon

118

binding in the EHBS will retain the capability to interact with Lys114, Lys125 and Arg129, and thus, turn the 'key' to open the 'lock'.

The design of sulfated flavonoids and sulfated tetrahydroisoquinoline derivatives was based on the structure of trisaccharide DEF and was time-consuming because it involved manual pre-positioning of the ligand in the pentasaccharide-binding site in AT. This reduces the structural variations that could be tested. In addition, the method involved a considerable degree of user-bias because binding modes are pre-defined. In order to increase our chances of targeting the PBS, we decided to use an approach that is solely based on the structure of the heparin-binding site (PBS + EHBS) in AT and allow for the selection of structures that favorably interact with the PBS. In addition, to explore a large number of possibilities, such a method would have to be rapid and automated. Given the success of our virtual library screening method with heparin/HS-protein interactions,[5] virtual screening seemed to be the most appropriate method to achieve the aforementioned goals.

7.2. Results and discussion

7.2.1. Virtual screening of non-saccharide sulfated small molecule libraries

Our studies on the docking of the H5 sequence and its variants to the activated crystal structure of AT revealed that the GOLD score of the H5 sequence is 148 and its geometry was repeatedly predicted to within 2.5 Å in multiple docking experiments (section 5.2.4). Thus, our search for small organic activators used these benchmarks as targets. We began with the bicyclic-unicyclic system present in IAS5, and 'synthesized' a

Bicyclic ring A   **Unicyclic ring B**

$S_1 \rightarrow S_9 = H/OSO_3^-$

$X = H/COO^-$

$L = -CO-$	$L = -CH_2CH_2CH_2CH_2-$
$L = -COO-$	$L = -CH_2CH_2CH_2CH_2CH_2-$
$L = -CONH-$	$L = -CH_2CH_2CH_2CH_2CH_2CH_2-$
$L = -CH_2-$	$L = -CH=CHCH_2CH_2-$
$L = -CH_2CH_2-$	$L = -CH_2CH=CHCH_2-$
$L = -CH_2CH_2CH_2-$	$L = -CH_2CH_2CH=CH-$
$L = -CH=CH-$	$L = -CH=CHCH_2CH_2CH_2-$
$L = -CH_2CH=CH-$	$L = -CH_2CH_2CH_2CH=CH-$
$L = -CH=CHCH_2-$	
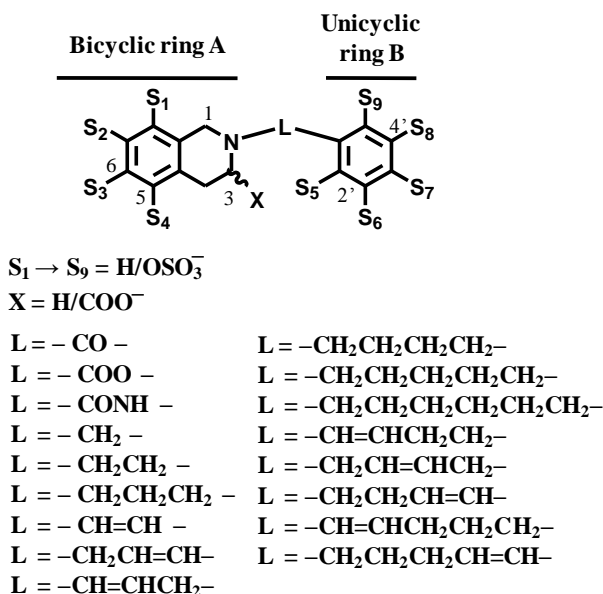
Figure 38. Combinatorial virtual library of bicyclic-unicyclic structures based on IAS5 with linkers containing 1-6 atoms. The library was created by variation of substituents $S_1 \rightarrow S_9$ and X, and the linker L.

virtual library of 24,576 chemical structures with 17 different linkers (L) containing 1-6 atoms and variations in the substitution pattern of sulfate groups ($S1 \rightarrow S8$, Figure 38). In addition, we decided to include both configurations of the carboxylate group at position 3 as well as molecules that are devoid of this group

The small molecule non-saccharide library was built using LEGION™, a powerful combinatorial library design software. LEGION™ rapidly generates 2D structures in Sybyl Line Notation in a combinatorial manner. A plausible 3D geometry of the 2D structures generated by LEGION was obtained using CONCORD™ and these structures were subject

to energy minimizations (see section 5.4.2) in an automated manner using in-house SPL scripts. The SPL scripts were written to simultaneously modify the atom and bond types of the sulfate groups so that they were identical to the types used for GAGs (section 5.4.4). The protocol used for virtual screening of non-saccharide molecules was essentially identical to that used for saccharides (Figure 33).

The dual filter (GOLD score + geometry convergence) selected 93 'hits' from the 24,576 structures in the virtual library. The GOLD scores of these hits ranged from 101 to 132 as compared to 148 and 122 for the H5 sequence and a hexasaccharide that contains the H5 sequence and an additional IdoAp2S residue at its non-reducing end.[*]
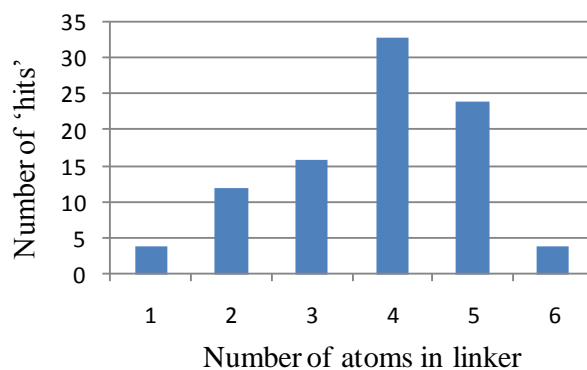
Figure 39. Histogram plot of the number of atoms in the linker for the 93 'hits' obtained from virtual screening of a non -saccharide sulfated library

Of the 93 'hits', 33 had an all carbon 4-atom linker, while 24 had an all carbon 5-atom linker (Figure 39). Of the 57 'hits' that had a 4-5 carbon linker, 38 had a trans

---

[*] When compared to H5, the hexasaccharide (H6) was found to consistently dock in a similar geometry although laterally displaced by a few Å. This resulted in a GOLD score of ~ 25 units less (section 5.2.6).

double bond with a strong preference to be alpha to either the nitrogen atom or the phenyl ring (Figure 38). Most of the remaining 19 of 57 'hits' contained a fully saturated methylene linker while a few contained a cis double bond.

With respect to number and distribution of negative charges, majority of 'hits' contained 4-6 negative charges. It is interesting that molecules possessing sulfation levels in excess of 6 were not selected to be 'hits'. Sixty nine of the 93 'hits' contained a

3,5,6-trisubstituted tetrahydroisoquinoline fragment
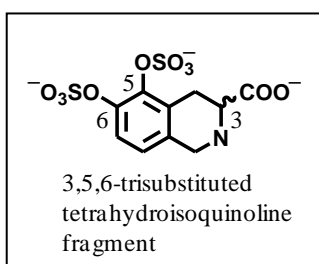
carboxylate group at the 3-position of the tetrahydroisoquinoline ring. Of these, 43 were the R isomer and 26 were the S isomer. Most interestingly, 48 of the 93 'hits' contained a 3,5,6-trisubstituted tetrahydroisoquinoline fragment (see figure)

suggesting that sulfate groups at the 5 and 6-position of the tetrahydroisoquinoline ring favorably contribute towards affinity.

The number and position of sulfate groups on the unicyclic ring of the 'hits' are variable and do not show clearly discernible patterns. In fact, the position of sulfate groups on the unicyclic ring appears to correlate with the length of the linker. For 'hits' with 4-5 carbon linkers, a sulfate group is located at the 2' position for majority of the compounds, while for 'hits' with shorter linkers, sulfate groups tend to be at the 3', 4' and/or 5' positions. Since the 'hits' with 4-5 carbon linkers are highest scoring, these results suggest that the sulfate group at the 2'position interacts with a basic amino acid in AT (see Table 5, below).

Table 5. Summary of structural features of most favorable 'hits'
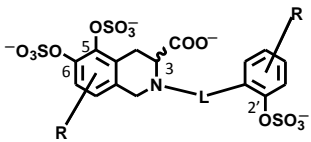
	Number of 'hits'	GOLD Scores	Percentage with 4 or 5 carbon linkers
	37	104-132	76%

Figure 40 shows an overlay between pentasaccharide $H5_{CRYS}$ (see Figure 24 for structure) and R-ACT6955, a representative 'hit' from the virtual screening study that contains the 3,5,6-trisubstituted tetrahydroisoquinoline fragment shown above and a 2'sulfate group in the unicyclic ring. The indicated pose of R-ACT6955 was repeatedly predicted in triplicate docking experiments to within 2.5 Å as a consequence of our dual-filter protocol. The figure shows striking similarities in the location of 4 negative charges. While the 3 carboxylate, 5-sulfate and 6-sulfate groups of R-ACT6955 overlap with the E-ring carboxylate, F-ring *N*-sulfate and G-ring carboxylate groups of $H5_{CRYS}$, the 2' sulfate group of R-ACT6955 overlaps with the D-ring 6-sulfate group of $H5_{CRYS}$. These charges interact with the critically important positive triad (Lys114, Lys125 and Arg129) which is located in the pentasaccharide-binding site (PBS) in antithrombin (AT). Additionally, the 5' sulfate group of R-ACT6955 interacts with the extended heparin-binding site residue Arg132. Thus, while the bicyclic ring binds right in the center of the PBS, the unicyclic ring binds at the interface of the PBS and EHBS. It is important to note that the 6-sulfate group in R-ACT6955 overlaps with a carboxylate group of $H5_{CRYS}$. Since sulfate groups contain three potential interaction points, the 6-sulfate group is predicted to simultaneously
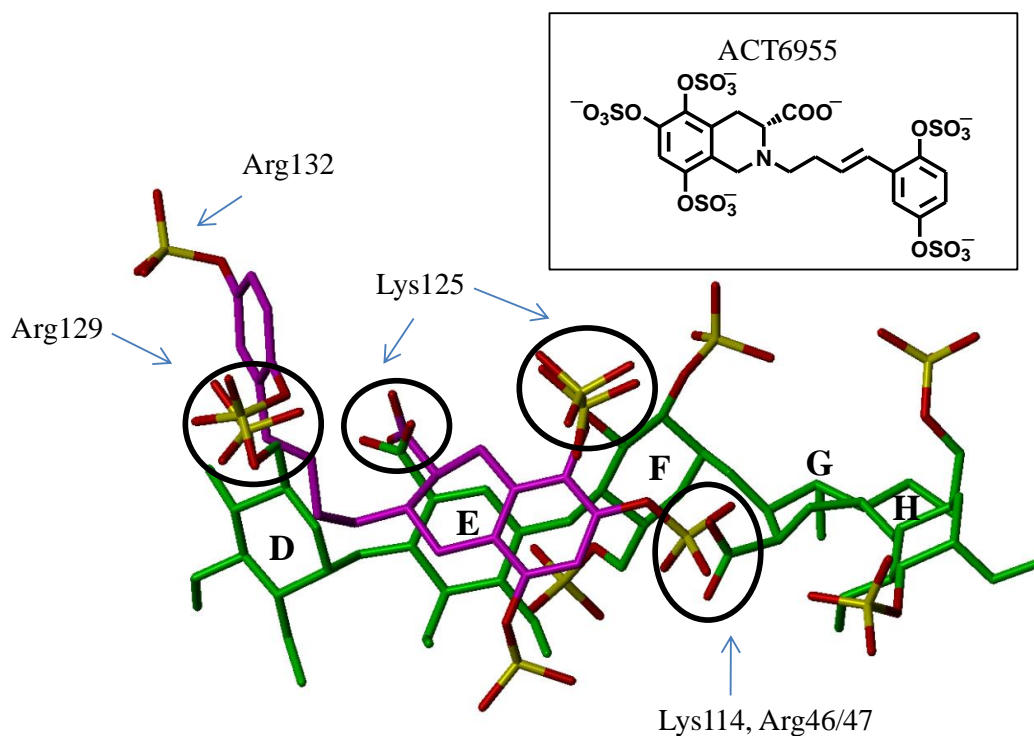
Figure 40. Overlay of heparin pentasaccharide H5 (green, obtained from the crystal structure 1TB6.pdb) and the repeatedly predicted docking pose of R-ACT6955 (magenta, structure shown in in-lay), a hit from the virtual screening study of sulfated non-saccharide molecules. Overlying negative charges (sulfate + carboxylate groups) are encircled (black). The critical PBS residues Lys114, Lys125 and Arg129 interact with the 6-sulfate, 5-sulfate and 2'-sulfate groups of R-ACT6955, respectively, which in turn overlay with the G-ring carboxylate, F-ring *N*-sulfate and D-ring 6-sulfate groups of the H5 sequence. In addition, the 3-carboxylate of R-ACT6955 and the E-ring carboxylate of H5 overlap and interact with Lys125. Note that R-ACT6955 is the R isomer.

interact with Lys114 (ionic) and the N-terminal residue Arg47 (hydrogen bond) whereas the G ring carboxylate of H5$_{CRYS}$ appears to interact only with N-terminal residue Arg46 (hydrogen bond). It is also important to note that the overlap of the E ring carboxylate and3-carboxylate of R-ACT6955 will be dependent on the configuration at the 3 position of the tetrahydroisoquinoline ring. For optimal overlap, the 'R' configuration is necessary.

The most appealing result of our virtual screening study is the dependency of the overlap shown in Figure 40 with linker length for the series of 'hits' that contain fragment I and the 2' sulfate group of the unicyclic ring. Figure 42 reveals that irrespective of the exact nature of the linker, the docking pose orients the interacting negative charges in an identical manner. For example, R-ACT6748 with a 5-carbon linker containing a cis-double bond has a binding mode similar to R-ACT6687 which contains a fully saturated linker. In fact, the unicyclic ring varies to position the 2' sulfate group identically. This binding mode does not depend on the configuration of the carboxylate group (R and S isomers dock similarly) or even the presence of this group (not shown). Thus, according to our results, the interaction between the 3-carboxylate group and Lys125 is a redundant interaction since S isomers cannot engage in a productive interaction. Nevertheless, structure-activity studies of pentasaccharide H5 using masked carboxylates (methyl esters) revealed that the E ring carboxylate plays an important role in activity.[*]

Based on the above results, it was decided to select R-ACT6955 as a synthetic target. R-ACT6955 possesses structural features that are representative of the highest scoring

---

[*] The authors measured activity using a factor Xa inhibition assay. Hence it is not known if the E ring carboxylate is important for antithrombin affinity or activation or both.

'hits'— a) A four carbon linker possessing a trans double bond, b) a 3,5,6-trisubstituted tetrahydroisoquinoline fragment, c) a unicyclic ring bearing sulfate groups at the 2' and 5' positions that could interact with Arg129 and Arg132, respectively, and d) R configuration at position 3 which allows for optimal overlap with the E-ring carboxyate group of the H5 sequence. Due to the structural complexity of R-ACT6955, S-ACT3227 was selected as a model compound because of the commercial availability of its tetrahydroisoquinoline fragment. In addition, the 'right wing' of S-ACT3227 is structurally identical to that of R-ACT6955. Thus, the series of reactions that are used to construct S-ACT3227 may also be used to synthesize R-ACT6955. Finally, S-ACT3227 is not a 'hit' and will serve as a negative control to our design strategy.
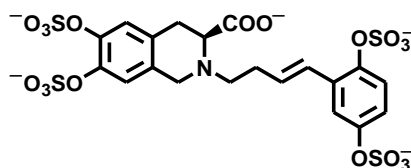
Figure 41. Structure of S-ACT3227

7.2.2. Synthesis of model compound S-ACT3227

7.2.2.1. Synthetic plan

We envisioned that S-ACT3227 can be synthesized from the key intermediate **1** in three steps namely demethylation using boron tribromide, microwave-assisted sulfation and selective ester cleavage (Figure 43). The conditions for these three steps have already
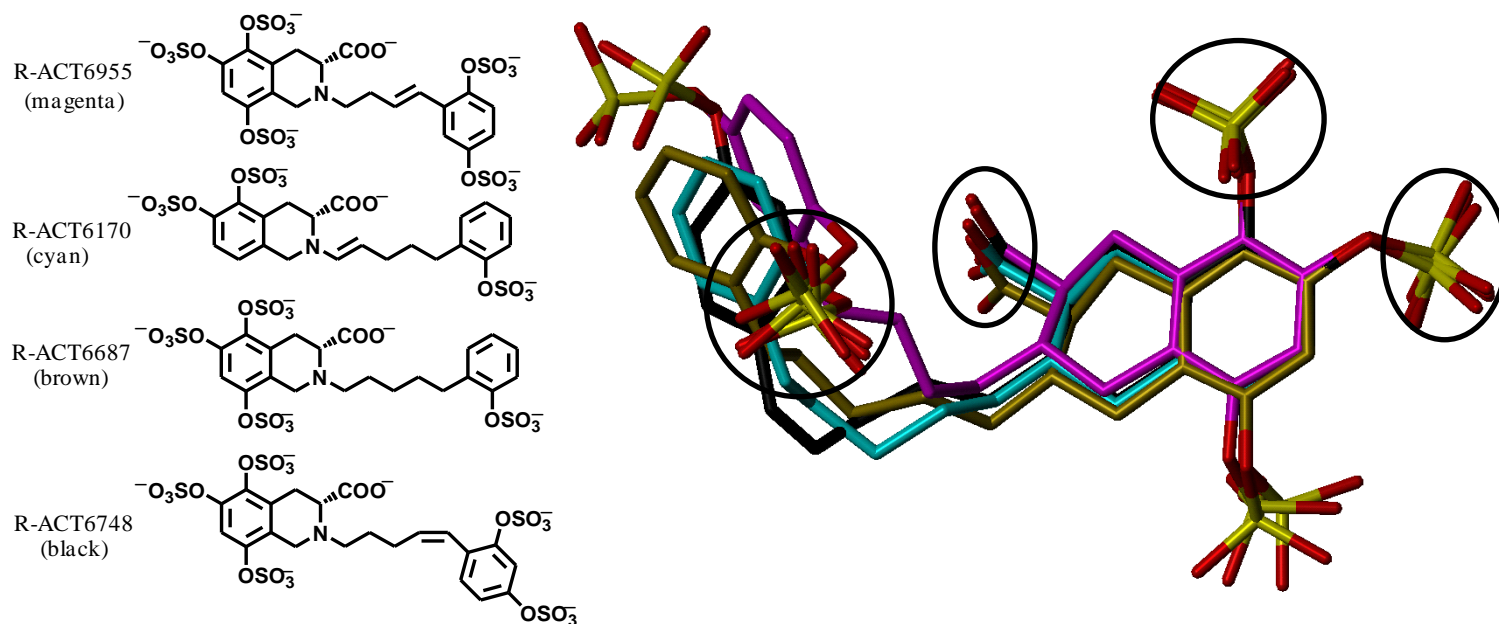
Figure 42. Overlay of different docked poses of 'hits' cotaining core fragment I and 2'-substituted unicyclic ring. Note the striking similarity of ortientation of the bicyclic ring and encircled negative charges.
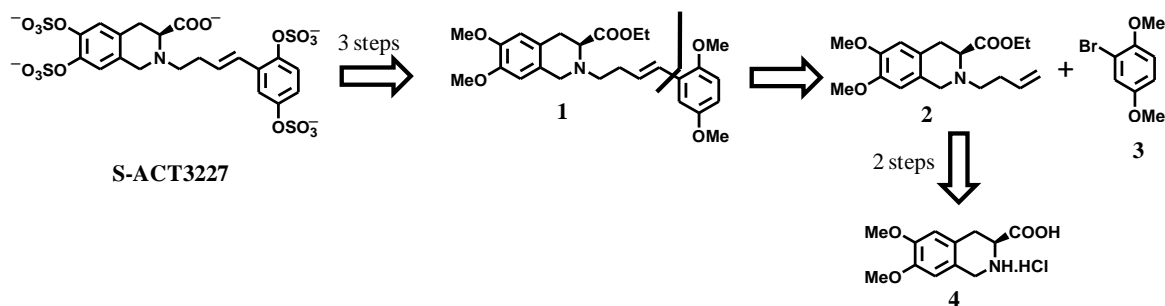
Figure 43. Retrosyntheic analysis of S-ACT3227

been worked out in our laboratory (chapters 2 and 3). Disconnection of the olefin-phenyl

ring linkage will result in amine **2** and substituted bromobenzene **3**. In the forward sense, a

palladium catalyzed Heck reaction of intermediates **2** and **3** will deliver **1**. Intermediate **2**

may be rapidly assembled from the commercially available substituted

tetrahydroisoquinoline **4**.

7.2.2.2. Synthesis of S-ACT3227

Starting with commercially available **4**, acid-catalyzed esterification in ethanol

solvent yielded **5** in 74 % yield (Figure 44A). Alkylation of the amine **5** using 4-bromo-1-

butene and diisopropylethylamine (DIEA) base resulted in the tertiary amine **2** in 83 %

yield. However, the Heck coupling of **2** and **3** proved to be problematic. Only traces of

product were detected when palladium (II) acetate (5 mol%) and either tris-4-

methoxyphenylphosphine, tri-o-tolylphosphine or triphenylphosphine (10-20 mol%) was

used as ligand in acetonitrile solvent. The yield of the reaction could not be improved in

DMF or dioxane. Since significant quantities of **3** could be isolated from the reaction

mixture, it was hypothesized that oxidative insertion of the Pd(II) catalyst was not taking place at significant rates. Thus, the electronic nature of the benzene ring was changed by replacing methyl groups with acetate groups (**3** →**7** →**8**) using standard protecting group manipulations (Figure 44B). While the coupling of **2** and **8** proved to be unproductive initially, the trans product (*J* = 15.9 Hz) was eventually isolated in 10 % yield when the reaction was carried using the Pd(0) catalyst tetrakistriphenylphosphine in acetonitrile solvent and potassium carbonate as base in the presence of 4 Å molecular sieves. Unfortunately, the yields were variable and were dependent on the scale of the reaction. Attempts to increase the yield of the reaction by adding stoichiometric amount of catalyst in 4 portions over a time interval of 24 hours only resulted in a decrease in yield. Gratifyingly, when the reaction was carried out under microwave conditions (140 ℃, 45 min) on a 50 mg scale (0.16 mmol), the yield improved to 37 %. However, microwave synthesis at larger scales (e.g. 500 mg) delivered only trace amount of product.

Settling for 5-20 % yields under non-microwave conditions, intermediate **9** was carried forward in multi-gram quantities. Thus, deprotection of the acetate groups using sodium bicarbonate in aqueous methanol at 50 ℃ resulted in **10**. Demethylation of **10** using boron tribromide yielded a yellow solid, which was subjected to microwave-based sulfation to yield the pentasulfated product **11**. The structure of **11**, which showed a single peak on capillary electrophoresis, was deduced from ^{1}H-NMR and 2D-HSQC experiments. Thus, under microwave conditions, the double bond gets sulfonated probably via an addition-elimination mechanism. For the purpose of screening, **11** (0.01 mmol, 11 mg) was treated with potassium *tert*-butoxide and water in dimethylsulfoxide solvent at room
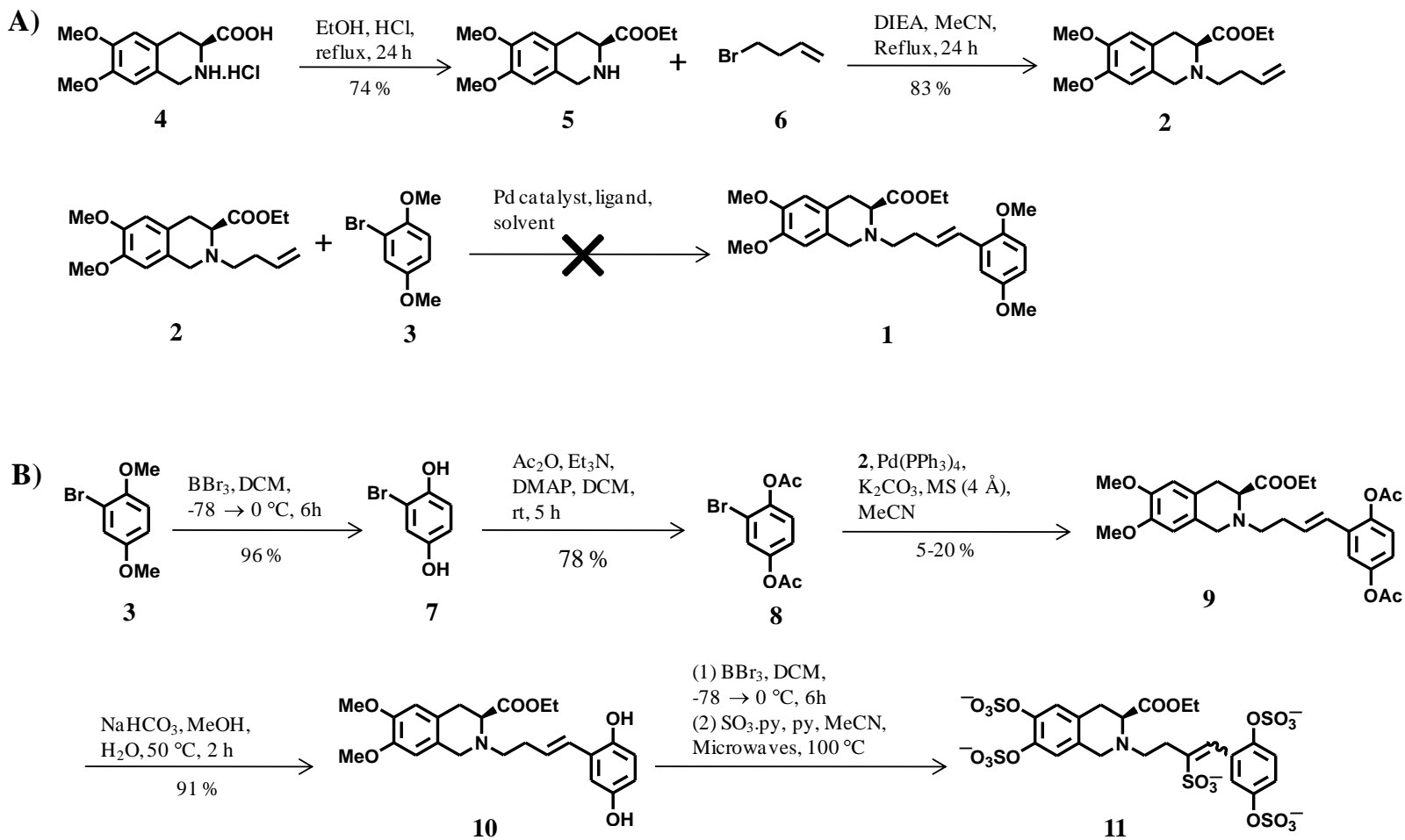
Figure 44. Synthetic efforts towards S-ACT3227 resulted in **11** containing an additional sulfonate group
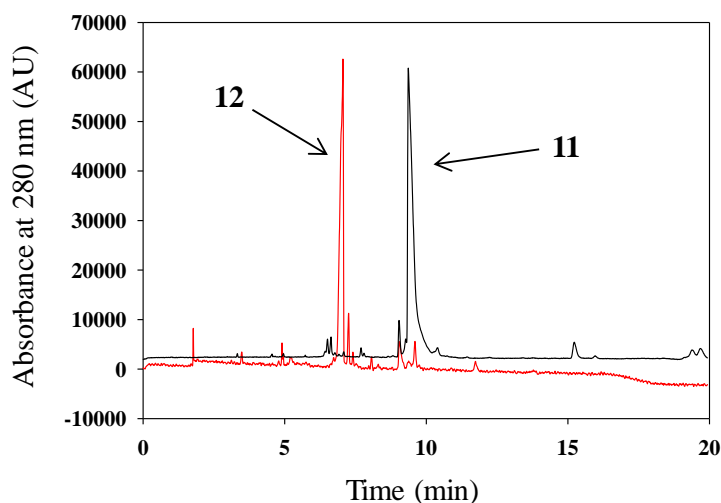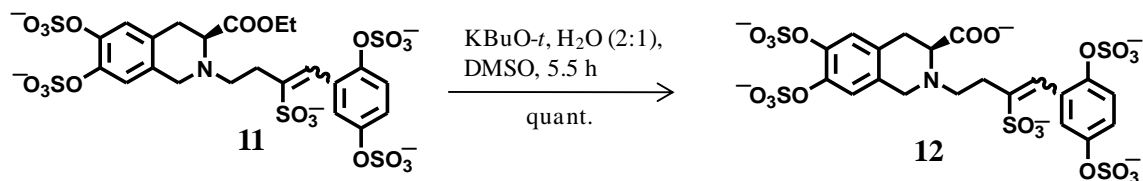
Figure 45. Capillary electrophoretic monitoring of the deprotection of ethyl ester in **11**. The analysis was carried out under reverse polarity conditions (-10 kV) so that the more negatively charged compounds elute earlier. At zero minutes, before the addition of potassium *tert*-butoxide, the black electropherogram was obtianed which corresponds to starting material **11**. The analysis revealed that the reaction was completer after 5.5 hours (red electropherogram).
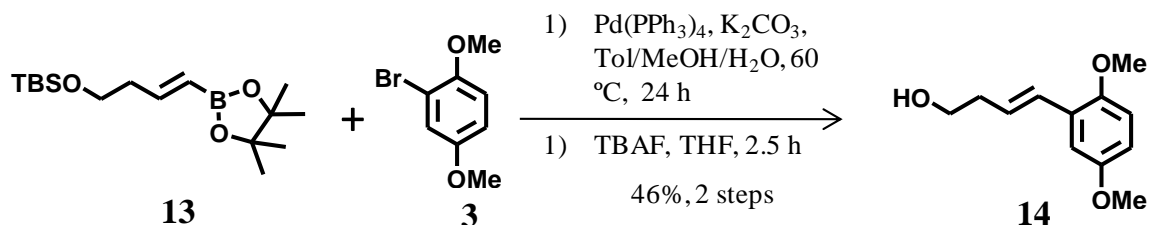
Figure 46. Efficient synthesis of the 'right wing' of activators R-ACT6955 and S-ACT3227.

temperature. The reaction was monitored by capillary electrophoresis (Figure 45) and the product (**12**) was isolated in essentially quantitative yield.

Biochemical evaluation of **11** and **12** revealed that both these compounds could not activate antithrombin up to a concentration of 250 μM (not shown). This suggests that these molecules do not bind either the PBS or the EHBS in antithrombin. While non-specific binding to the EHBS has been eliminated, it is hoped that appropriate virtual-screening-guided structural modifications may induce specific binding to the desired PBS.

At this stage, it was decided to shift focus and devise a strategy to efficiently synthesize the 'right wing' of R-ACT6955/S-ACT3227. The Heck reaction under optimized conditions (Pd(Ph$_3$)$_4$, K$_2$CO$_3$, MS (4 Å)) delivered low isolated yields of product because of a significant amount of mixed fraction obtained from silica gel column chromatography containing **9** and the *exo*-methylene regioisomer (Figure 44B). In fact only ~ 5% of bromobenzene **8** could be recovered from the reaction mixture. This led us to explore an alternative coupling strategy that would exclusively deliver the trans product.
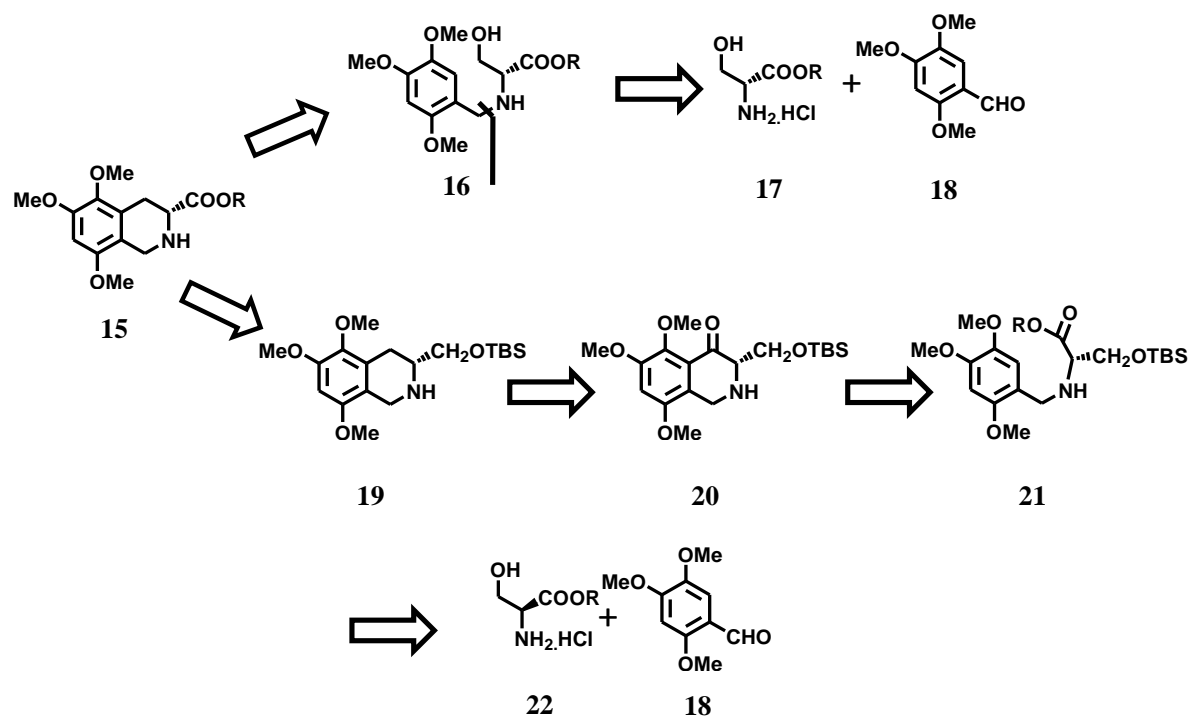
Figure 47. Retrosynthetic analysis of substituted tetrahydroisoquinoline **15**

Our efforts led to the synthesis of the trans product **14** in moderate yields by a Suzuki coupling of the pinacolboronate **13** and bromobenzene **3** without the need of using acetate protecting groups (Figure 46). **14** may then be used to generate suitable coupling partners for the synthesis of the carbon skeleton of potential activators, thus making the overall synthesis highly convergent. Given that the double bond gets functionalized under microwave-assisted sulfation conditions, it was decided that **14** needs to be hydrogenated before coupling in the course of future studies.

7.2.3. Synthetic efforts towards the substituted tetrahydroisoquinoline fragment in R-ACT6955

7.2.3.1. Synthetic plan

The presence of a chiral center in substituted tetrahydroisoquinoline **15** was a major factor in deciding the synthetic route (Figure 47). Recognizing that **15** contains an α-amino carboxylic acid moiety, we decided to use disconnections that would lead us to the aminoacid serine which is commercially available in optically pure form. Disconnections in such a manner would obviate the need for asymmetric synthesis. Given that a host of differentially methoxy-substituted benzaldehydes are commercially available, we decided to use reductive aminations to construct the indicated C-N bond in **16**. Keeping these factors in mind, two different plans were conceived for the synthesis of substituted tetrahydroisoquinoline **15**. In plan A, **15** could be made by an intramolecular Friedel-Crafts alkylation reaction of **16**, which may be readily assembled from the D-serine derivative **17**

and aldehyde **18**. In plan B, a functional group transformation of the carboxylate ester in **15** to a primary alcohol followed by oxidative state adjustment would lead to key intermediate **20**. In the forward sense, **20** may be synthesized using the Friedel-Crafts acylation reaction from L-serine (**22**) and aldehyde **18**. Note that **17** and **22** are enantiomers.

7.2.3.2. Synthesis of substituted tetrahydroisoquinoline fragment in R-ACT6955

The execution of plan A was performed as outlined in Figure 48A. Aldehyde **23** and aminoacid **17** were treated with sodium triacetoxy borohydride in the presence of triethylamine in 1,2-dichloroethane to obtain **24** in 65 % yield. However, attempts to cyclize **24** using trifluorormethane sulfonic acid to generate the superelectrophile intermediate were unsuccessful.[234] Additionally, cyclization in acetic acid solvent or by using $BF_3.Et_2O$ following literature precedent failed.[235] To proceed further, it was decided to reduce the polarity of **24** by alkylation. Thus, treatment of **24** with 4-bromo-1-butene under basic conditions resulted in **25**. Based on literature precedent, it was hoped that oxidation of **25** under acidic oxidation conditions will generate the aldehyde which may spontaneously cyclize to generate the hydroxyl ester **27**.[236] Thus, **25** was subjected to Swern oxidation conditions to obtain a product whose structure could not be unambiguously assigned through NMR spectroscopy. When the product from the Swern
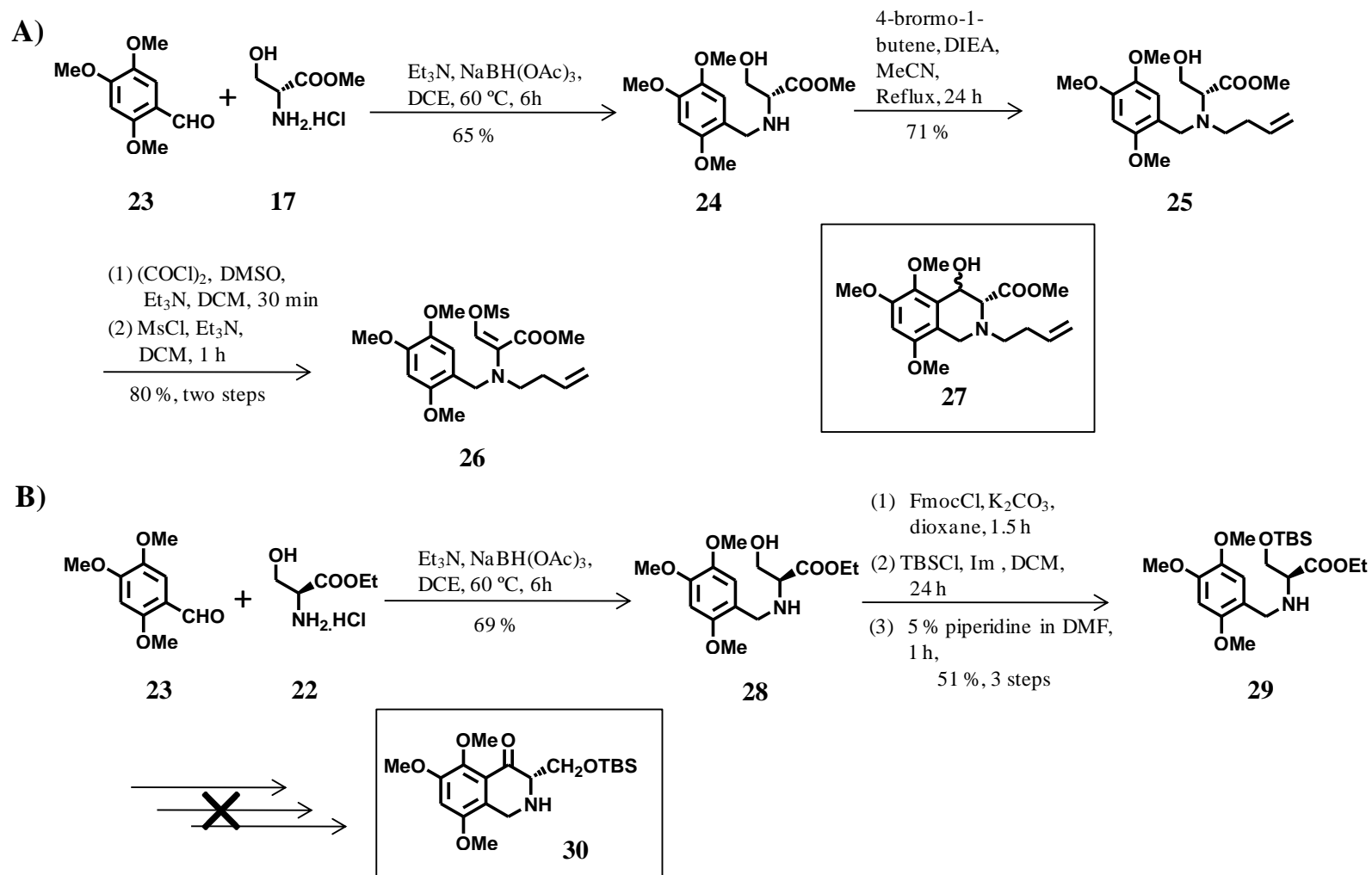
Figure 48. Synthetic efforts towards the substituted tetrahydroisoquinoline ring system present in R-ACT6955

oxidation of **25** was treated with mesyl chloride under basic conditions, the mesyl derivative **26** was formed. Thus, the product obtained from the oxidation of **25** was an equilibrium mixture of the aldehyde and the enol tautomer which was intractable to NMR analysis. Although the α,β-unsaturated compound **26** may be cyclized by treatment with a Lewis acid,[237,238] further work with this intermediate was not pursued because the configuration at C-3 would not be retained.

So far as plan B was concerned, reductive amination of aldehyde **23** using L-glycine ethyl ester **22** resulted in **28** (Figure 48B). Using standard protecting group chemistry, **28** was transformed to **29** in three high yielding steps. However, attempts to cyclize **29** via an aminoacid intermediate failed. Attempts to cyclize a simplified *tert*-butyl analog of **29** in which the CH_2OTBS group was absent (prepared via a different synthetic route from glycine as starting material, not shown) using trifluoroacetic acid and trifluoroacetic anhydride[239,240] were also unproductive. The reaction resulted in an unprecedented product whose structure could not be elucidated from ^{1}H and ^{13}C-NMR spectra.

7.3. Summary and future directions

Several promising 'hits' emerged from our virtual screening study of 27,576 sulfated, non-saccharide compounds with GOLD scores ranging from 101 to 132. A significant number of high scoring 'hits' contained a common 3,5,6-trisubstituted tetrahydroisoquinoline fragment. The optimal linkers contained 4-5 carbon atoms. Comparison of the binding modes of the top scoring 'hits' with that of the heparin pentasaccharide ($H5_{CRYS}$) revealed

key points of overlap between negative charges in the non-saccharide molecules with those in $H5_{CRYS}$. Additionally, the binding modes of these 'hits' are similar irrespective of the exact nature of the linker. These results are summarized in Table 5.

Two targets – R-ACT6955 and S-ACT3227 were selected for synthesis. Initial synthetic efforts led to the realization that the double bond in S-ACT3227 is incompatible with the microwave-assisted sulfation reaction. In addition, construction of an optically pure tetrahydroisoquinoline ring by incorporation of chirality inherent in L or D-serine proved to be problematic. A highly convergent synthetic approach was devised based on a Suzuki coupling reaction to afford the synthon containing a 4-carbon linker and the unicyclic ring. Coupling of this synthon or a derivative with the tetrahydroisoquinoline moiety will result in the carbon skeleton of the series of activators that contain a 4-carbon linker.

For future work, we will target racemates as this will permit cyclization to be conducted under the well-established Pictet-Spengler conditions.[241] Our plan is outlined in Figure 49. Starting from aldehyde **31**, condensation with nitromethane followed by lithium aluminium hydride (LAH) reduction will yield **33**, which has been reported in the literature.[242] Pictet-Spengler cyclization of **33** using formaldehyde will result in tetrahydroisoquinoline **34**. While **34** has not been reported to date, Pictet-Spengler cyclization had been reported on an analog of **33** containing a different methoxy substitution pattern on the phenyl ring.[243]

Aldehyde **31** will also be used to synthesize the 3-substituted tetrahydroisoquinoline **37**. Briefly, condensation with 2-nitroethanol will yield **35**, which

has been reported.[244] Silyl protection of the primary alcohol followed by LAH reduction and Pictet-Spengler condensation will result in **37**. Oxidation state adjustment of **37** followed by coupling with a hydrogenated derivative of **34** (Figure 46) will result in the carbon skeleton of potential activators.
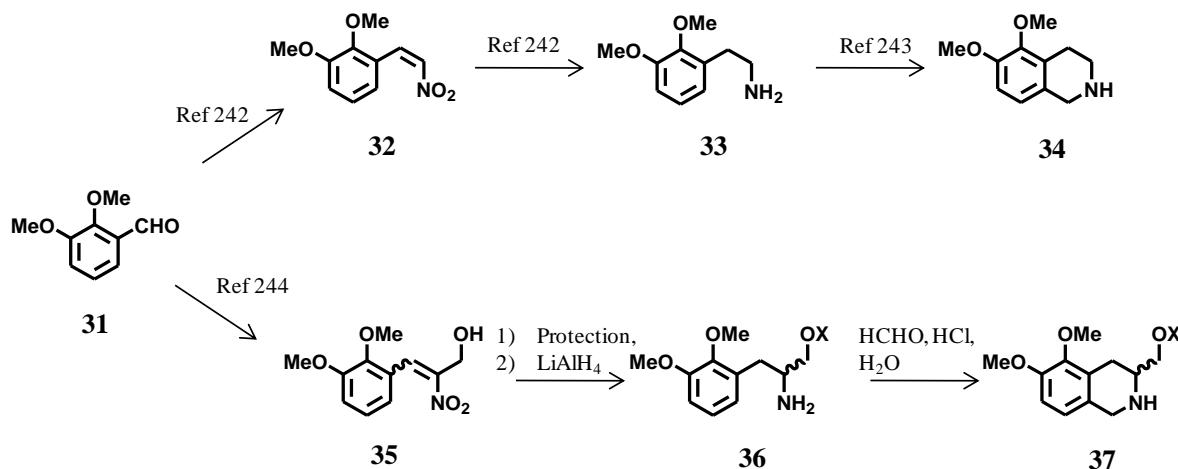
Figure 49. Re-visiting the synthesis of the substituted tetrahydroisoquinoline moiety present in majority of virtual screening 'hits'. The scheme is based on Pictet-Spengler cyclization of primary amines and will result in a racemate.

7.4. Experimental Section

2-But-3-enyl-6,7-dimethoxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid ethyl ester (2). To a stirred solution of 6,7-Dimethoxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid ethyl ester **5** (2.15 g, 8.1 mmol) in acetonitrile (27.0 mL) was added diisopropylethylamine (2.8 mL, 16.2 mmol) and 4-bromo-1-butene (1.6 mL, 16.2 mmol). The reaction was brought to reflux for 24 hours. After concentration, the residue was

purified by silica gel chromatography (4:1 → 3:1, hexanes:EtOAc) to afford **2** (2.14 g, 83 %) as a pale yellow oil. ^{1}H-NMR (CDCl$_3$, 400 MHz): δ 6.53 (s, 1H), 6.48 (s, 1H), 5.74-5.84 (m, 1H), 5.07-4.95 (m, 2H), 4.14-4.06 (2H, m), 3.95 (d, J = 15.2 Hz, 1H), 3.39 (s, 3H), 3.78 (s, 3H), 3.79-3.75 (d, J = 15.2 Hz, 1H), 3.70 (app t, J = 4.8 Hz, 1 H), 3.00 (dd, J = 16.0, 4.4 Hz, 2H), 2.82-2.68 (m, 2H), 2.33-2.28 (m, 2H), 1.19 (t, J = 7.2 Hz, 3H); ^{13}C-NMR (CDCl$_3$, 75 MHz): δ 172.8, 147.6, 136.7, 126.0, 124.2, 116.0, 111.3, 109.3, 60.6, 60.1, 56.07, 56.05, 54.4, 51.3, 32.5, 31.0, 14.6; ESI MS Calcd for C$_{18}$H$_{25}$NO$_4$ [M+H]$^+$ 320.19, Found 320.36

2-Bromo-benzene-1,4-diol (7). To a stirred solution of 2-Bromo-1,4-dimethoxy-benzene (2.50 g, 11.5 mmol) in dichloromethane (58.0 mL) was added a solution of boron tribromide (28.0 mL of a 1M solution in dichloromethane, 28.0 mmol) at -78 ºC. The resulting mixture was allowed to warm to room temperature. After 6 hours, the reaction mixture was cooled to 0 ºC and water (5 mL) was added dropwise. Ethyl acetate (200 mL) was added to the resulting suspension and vigorously stirred. After 30 minutes the suspension was filtered through a pad of Celite. A saturated solution of NH$_4$Cl (50 mL) was added to the filtrate and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (4 × 50 mL) and the combined layers were dried over Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was purified by silica gel chromatography (4:1, hexanes, EtOAc) to afford diol **7** (2.08 g, 96 %). ^{1}H-NMR (CD$_3$SOCD$_3$, 300 MHz): δ 9.35 (s, 1H), 9.04 (s, 1H), 6.84 (d, J = 3.0 Hz, 1H), 6.75 (d, J = 8.7 Hz, 1H), 6.58 (dd, J = 8.7, 2.4 Hz, 1H); ^{13}C-NMR (CD$_3$SOCD$_3$, 75 MHz): 151.1, 147.3, 119.5, 117.5, 116.1, 109.6

Acetic acid 4-acetoxy-2-bromo-phenyl ester (8). To a stirred solution of 2,2-Bromo-benzene-1,4-diol **3** (2.08 g, 11.1 mmol) in N.N-Dimethylformamide (13 mL) was added acetic anhydride (21 mL, 222 mmol), triethylamine (31 mL, 222 mmol) and DMAP (271 mg, 2.22 mmol). The reaction flask was rinsed with N.N-Dimethylformamide (2 mL) and stirred for 5 hours. Water (75 mL) was added to the reaction mixture at 0 ºC and stirred for 0.5 hours after which diethylether (150 mL) was added. The organic layer was separated and the aqueous layer was extracted with diethylether (4×100 mL). The combined layers were dried (Na_2SO_4) and concentrated to obtain a residue that was purified using silica gel chromatography (17:3, hexanes, EtOAc) to afford the diacetate **8** (2.35 g, 78 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): δ 7.39 (d, J = 2.4 Hz, 1H). 7.15-7.06 (m, 2H), 2.35 (s, 3H), 2.29 (s, 3H); ^{13}C-NMR (CDCl$_3$, 75 MHz): 169.1, 168.7, 148.7, 146.1, 126.7, 124.2, 121.9, 21.2, 21.0; ESI MS Calcd for $C_{10}H_9BrO_4$ [M+H]$^+$ 272.98, Found 273.01

2-[4-(2,5-Diacetoxy-phenyl)-but-3-enyl]-6,7-dimethoxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid ethyl ester (9). To a stirred solution of diacetate **8** (1.75 g, 6.4 mmol) and amine **2** (2.25 g, 7.0 mmol) in acetonitrile (128 mL) was added K_2CO_3 (2.65 g, 19.2 mmol), molecular sieves (4 Å, 500 mg) and tetrakispalladiumtriphenylphosphine (740 mg, 0.64 mmol). The reaction mixture was refluxed for 24 hours and filtered through a short pad of Celite. The filtrate was concentrated and purified by silica gel chromatography (3:2 →11:9, hexanes:EtOAc) to afford **9** (180 mg, 6 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): δ 7.21 (d, J = 2.7 Hz, 1H), 7.00 (d, J = 9 Hz, 1H), 6.93 (dd, J = 9, 2.7 Hz, 1H), 6.57 (s, 1H), 6.53 (s, 1H), 6.42 (d, J = 15.9 Hz,

1H), 6.23 (dt, J = 15.9, 6.6 Hz, 1H), 4.12 (m, 2H), 3.99 (d, J = 15.0 Hz, 1H), 3.84 (d, J = 15.0 Hz, 1H), 3.83 (s, 3H), 3.82 (s, 3H), 3.74 (dd, J = 6,4.8 Hz, 1H), 3.07 (dd, J = 12.0, 4.7 Hz, 1H), 2.96 (dd, J = 12, 3.3 Hz, 1H), 2.85 (m, 2H), 2.49 (m, 2H), 2.28 (s, 3H), 2.16 (s, 3H), 1.21 (t, J = 6.0 Hz, 3H); ^{13}C-NMR (CDCl$_3$, 75 MHz): δ 172.8, 169.6, 169.5, 148.6, 147.7, 145.2, 132.3, 131.6, 126.0, 124.2, 123.9, 123.6, 121.1, 119.5, 111.3, 109.3; ESI MS Calcd for C$_{28}$H$_{33}$NO$_8$ [M+H]$^+$ 512.23, Found 512.32

2-[4-(2,5-Dihydroxy-phenyl)-but-3-enyl]-6,7-dimethoxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid ethyl ester (10). To a stirred solution of 9 (180 mg, 0.35 mmol) in methanol (5 mL) and water (2.5 mL) was added a saturated solution of NaHCO$_3$ (2.5 mL) and gently heated to 50 ºC. After 4 hours, the reaction mixture was partitioned between brine and ethyl acetate. The organic layer was separated and the aqueous layer was extracted with ethyl acetate. The combined layers were concentrated *in vacuo* and purified by silica gel chromatography (1:1, hexanes, EtOAc) to afford **10** (136 mg, 91 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): δ 6.74 (d, J = 3.0 Hz, 1H), 6.63 − 6.52 (m, 5H), 6.10 (dt, J = 16.2, 6.6 Hz, 1H), 4.11 − 4.18 (m, 2H), 4.02 (d, J = 15.3 Hz, 1H), 3.82 (d, J = 15.3 Hz, 1H), 3.83 (s, 3H), 3.82 (s, 3H), 3.60 (app t, J = 6.0 Hz, 1H), 3.10 (dd, J = 18, 5.6 Hz, 1H), 3.02 (dd, J = 15.0, 4.8 Hz, 1H), 2.93 − 2.79 (m, 2H), 2.53 − 2.46 (m, 2H), 1.22 (t, J = 7.2 Hz, 3H); ESI MS Calcd for C$_{24}$H$_{29}$NO$_6$ [M+H]$^+$ 428.21, Found 428.40

2-[4-(2,5-Dihydroxy-phenyl)-3-hydroxy-but-3-enyl]-6,7-dihydroxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid ethyl ester pentasulfate sodium salt (11). To a stirred solution of **10** (136 mg, 0.32 mmol) in dichloromethane (6.4 mL) was added a solution of boron tribromide (1.9 mL of a 1M solution in dichloromethane, 1.9 mmol) at -

78 ºC. The resulting mixture was allowed to warm to room temperature. After 6 hours, the reaction mixture was cooled to 0 ºC and water (1 mL) was added dropwise. Ethyl acetate (50 mL) was added to the resulting suspension and vigorously stirred. After 30 minutes the suspension was filtered through a pad of Celite. A saturated solution of NH_4Cl (10 mL) was added to the filtrate and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (4 × 30 mL) and the combined layers were dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by passing through a short plug of silica gel (1:4, hexanes:EtOAc) to afford a yellow solid (22 mg). To a stirred solution of the resulting yellow solid in acetonitrile (500 μL) and pyridine (194 μL) was added pyridine-sulfur trioxide complex (350 mg). The resulting mixture was heated in a CEM-Discover microwave synthesizer at 100 ºC for 30 minutes under sealed conditions. On cooling, the reaction was quenched with methanol (2 mL), concentrated and chromatographed on a Sephadex G10 column (75 × 1.5 cm) with water as eluant. Appropriate fractions (capillary electrophoretic analysis) were pooled, concentrated in vacuo. The sodium salt of the title compound (**11**) was obtained by chromatography using SP-Sephadex C25 sodium exchanger followed by lyophilization (11 mg, 4 %, 2 steps).^{1}H-NMR (D_2O, 400 MHz): δ 7.24 (br s, 1H), 7.22 (s, 1H), 7.19 (s, 1H), 7.13 (d, J = 2.8 Hz, 1H), 7.10 (dd, J = 8.8, 2.8 Hz, 1H), 7.02 (d, J = 8.8 Hz, 1H), 4.22 (d, J = 14.8 Hz, 1H), 4.14 (t, J = 4.5 Hz, 1H), 4.08 – 4.01 (m, 3H), 3.15 – 3.11 (m, 2H), 3.03 (app t, J = 6.8 Hz, 2H), 2.61 (app t, J = 6.4 Hz), 1.03 (t, J = 7.2 Hz, 3H); 2D-HSQC (D_2O, 400 MHz, ^{1}H; 100 MHz , ^{13}C): δ (^{1}H, ^{13}C) 7.22, 143.5; 7.22, 122.8; 7.10, 122.1; 7.19, 121.5; 7.13, 120.8;

7.03, 118.8; 4.04, 63.2; 4.14, 60.0; 3.03, 53.6; 4.07, 50.8; 4.20, 50.8; 2.61, 29.5; 3.16, 27.9; 1.03, 13.4

2-[4-(2,5-Dihydroxy-phenyl)-3-hydroxy-but-3-enyl]-6,7-dihydroxy-1,2,3,4-tetrahydro-isoquinoline-3-carboxylic acid pentasulfate sodium salt (12). To a stirred solution of **11** (10 mg, 0.012 mmol) in dimethylsulfoxide (120 μL) was added sodium *tert*-butoxide (3.5 mg, 0.036 mmol) and water (0.3 μL). After 5.5 hours, a solution of Na$_2$HPO$_4$ (3.4 mg) in water (1 mL) was added. After 30 minutes, the reaction mixture was directly loaded onto a Sephadex G10 column and eluted with water. Appropriate fractions were pooled (capillary electrophoresis) and lyophilized to obtain the title compound **12** (10 mg, quant.). ^{1}H-NMR (D$_2$O, 300 MHz): δ 7.42 (s, 1H), 7.29 – 7.10 (m, 5H), 4.55 (d, J = 15.8 Hz, 1H), 4.34 - 4.20 (m, 3H), 3.37 - 3.21 (m, 2H), 3.05 – 2.96 (m, 2H), 2.78 – 2.71 (m, 2H).

3-Hydroxy-2-(2,4,5-trimethoxy-benzylamino)-propionic acid methyl ester (24). To a stirred solution of 2,4,5-Trimethoxy-benzaldehyde **23** (3.4 g, 17.4 mmol) and D-Serine hydrochloride methyl ester **17** (2.7 g, 17.4 mmol) in 1,2-Dichloro-ethane (62 mL) was added trithethylamine (4.9 mL). After 10 minutes, Sodium triacetoxyborohydride (5.2 g, 24.4 mmol) was added portion-wise. The reaction mixture was heated to 60 ℃ for 6 hours. Ethyl acetate (150 mL) and saturated sodium bicarbonate (75 mL) was added to the reaction mixture and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (3 × 100 mL) and the combined layers were concentrated *in vacuo*. The residue was purified by silica gel chromatography (99:1, EtOAc:MeOH) to afford the title compound **24** (3.36 g, 65 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): δ 6.76 (s, 1H), 6.47 (s, 1H),

3.83 (s, 3H), 3.79 (s, 3H), 3.75 – 3.69 (m, 3H), 3.64 (s, 3H), 3.58 (dd, J = 10.8, 6.3 Hz, 1H), 3.35 (app dd, J = 6.3, 4.5 Hz, 1H). ^{13}C-NMR (CDCl$_3$, 75 MHz): δ 173.7, 152.2, 149.1, 142.9, 119.0, 97.6, 62.4, 62.0, 56.8, 56.39, 56.37, 52.2, 47.1, ESI MS Calcd for C$_{14}$H$_{21}$NO$_6$ [M+H]$^+$ 300.14, Found 300.18

2-[But-3-enyl-(2,4,5-trimethoxy-benzyl)-amino]-3-hydroxy-propionic acid methyl ester (25). To a stirred solution of **24** (300 mg, 1.0 mmol) in acetonitrile (2 mL) and diisopropylethylamine (0.9 mL) was added 4-Bromo-1-butene (500 μL, 5.0 mmol). The reaction mixture was microwaved at 150 ºC for 20 minutes under sealed conditions. The reaction mixture was concentrated in vacuo and purified by flash chromatography on SiO$_2$ (3:2, hexanes:EtOAc) to afford the title compound **25**. ^{1}H-NMR (CDCl$_3$, 300 MHz): 6.78 (s, 1H), 6.50 (s, 1H), 5.70 – 5.56 (m ,1H), 5.00 – 4.92 (m, 2H), 3.91 (d, J = 12.9 Hz, 1H), 3.87 (s, 3H), 3,81 (s, 3H), 3.79 (s, 3H), 3.72 (s, 3H), 3.70 – 3.65 (m, 2H), 3.62 – 3.53 (m, 2H), 3.28 (br s, 1H), 2.79 – 2.61 (m, 2H), 2.17 – 2.09 (m, 2H), 1.78 (br s, 1H); ^{13}C-NMR (CDCl$_3$, 75 MHz): 172.0, 152.5, 149.2, 142.8, 137.0, 118.1, 116.3, 115.6, 97.6, 63.7, 58.9, 56.9, 56.3, 56.2, 51.5, 50.4, 50.3, 33.1, ESI MS Calcd for C$_{18}$H$_{27}$NO$_6$ [M+H]$^+$ 354.19, Found 354.30

2-[But-3-enyl-(2,4,5-trimethoxy-benzyl)-amino]-3-methanesulfonyloxy-acrylic acid methyl ester (26). To a stirred solution of oxalyl chloride (37.5 μL, 0.44 mmol) in dichloromethane (1.0 mL) was carefully added dimethylsulfoxide (62.5 μL, 0.88 mmol) at -78 ºC. After 10 minutes, a solution of **25** (76 mg, 0.22 mmol) in dichloromethane (0.7 mL) was added. After 15 minutes, Et$_3$N (123 μL, 0.88 mmol) was carefully added. After

15 minutes, the reaction was slowly allowed to warm to room temperature. After 30 minutes, the reaction mixture was placed on a silica gel column and the product was isolated by flash chromatography using hexanes and EtOAc (1:4) to afford the aldehyde (74 mg, 95 %). To a stirred solution of the resulting residue (35 mg, 0.12 mmol) in dichloromethane (400 µL) and Et$_3$N (25 µL) was added methanesulfonyl chloride (10.3 µL) at 0 °C. After 1 hour, the reaction mixture was diluted with dichloromethane (2 mL) and loaded onto a short column of silica gel. The title compound (**26**) was isolated by flash chromatography using hexanes and ethyl acetate (13:7) (43 mg, 84 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): 7.19 (s, 1H), 6.98 (s, 1H), 6.48 (s, 1H), 5.85 – 5.72 (m, 1H), 5.07 – 4.96 (m, 2H), 4.14 (br s, 2H), 3.86 (s, 3H), 3.83 (s, 3H), 3.77 (s, 3H), 3.76 (s, 3H), 3.07 (m, 2H), 2.88 (s, 3H), 2.22 (m, 2H); ^{13}C-NMR (CDCl$_3$, 75 MHz): 165.8, 151.9, 148.8, 143.1, 136.3, 135.0, 128.4, 118.5, 116.3, 113.9, 97.6, 56.8, 56.5, 56.4, 52.3, 51.9, 50.3, 37.9, 33.0; ESI MS Calcd for C$_{19}$H$_{27}$NO$_8$S [M+H]$^+$ 430.15, Found 430.24

3-Hydroxy-2-(2,4,5-trimethoxy-benzylamino)-propionic acid ethyl ester (28).
To a stirred solution of 2,4,5-Trimethoxy-benzaldehyde **23** (2.00 g, 10.2 mmol) and L-Serine ethyl ester **22** (1.73 g, 10.2 mmol) in 1,2-Dichloro-ethane (37 mL) was added Et$_3$N (2.9 mL, 20.4 mmol). After 10 minutes, Sodium triacetoxyborohydride (3.03 g, 14.3 mmol) was added and the reaction mixture was heated to 60 °C. After 6 hours, ethyl acetate (100 mL) and saturated sodium bicarbonate (50 mL) was added to the reaction mixture and the organic layer was separated. The aqueous layer was extracted with ethyl acetate (3 × 75 mL) and the combined layers were concentrated *in vacuo*. The residue was purified by silica gel chromatography to afford the title compound (2.2 g, 69 %). ^{1}H-NMR

(CDCl$_3$, 300 MHz): δ 6.78 (s, 1H), 6.50 (s, 1H), 4.12 (q, $J = 7.1$ Hz, 2H), 3.86 (s, 3H), 3.82 (s, 3H), 3.80 (s, 3H), 3.78 – 3.73 (m, 3H), 3.59 (dd, $J = 9, 6.2$ Hz, 1H), 3.36 (dd, $J = 6.0, 4.5$ Hz, 1H), 1.13 (t, $J = 7.1$ Hz, 3H)); ^{13}C-NMR (CDCl$_3$, 75 MHz): 173.3, 152.2, 149.1, 142.9, 119.1, 114.3, 97.6, 62.4, 62.0, 61.3, 56.8, 56.41, 56.39, 47.2, 14.4

3-(tert-Butyl-dimethyl-silanyloxy)-2-(2,4,5-trimethoxy-benzylamino)-propionic acid ethyl ester (29). To a stirred solution of **28** (2.44 g, 7.8 mmol) in 1,4-Dioxane (31 mL) was added K$_2$CO$_3$ (2.20 g, 15.6 mmol) and Fmoc-Cl (2.12 g, 8.2 mmol). After 1.5 hours, the reaction mixture was partitioned between ethyl acetate (150 mL) and water (50 mL). The organic layer was separated and the aqueous layer was extracted with ethyl acetate (3 × 100 mL). The combined layers were concentrated and the Fmoc derivative was isolated by flash chromatography on SiO$_2$ (1:1, hexanes:EtOAc) (3.52 g, 84 %). The resulting residue was dissolved in dichloromethane (66 mL) and treated with imidazole (899 mg, 13.2 mmol) and TBS-Cl (995 mg, 6.6 mmol). After 24 hours, saturated NH$_4$Cl (50 mL) and dichloromethane (50 mL) were added. The organic layer was separated and the aqueous layer was extracted with dichloromethane (3 × 75 mL). The combined layers were concentrated *in vacuo*. The crude residue obtained (3.4 g) was dissolved in N,N-dimethyl formamide (10 mL) and piperidine (0.5 mL) was added. After 1 hour, the reaction mixture was diluted with toluene (50 mL) and concentrated in vacuo. This procedure was repeated three times to obtain a solid which was purified by flash chromatography (7:3, hexanes:EtOAc) to obtain the title compound (**29**) (1.7 g, 76 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): 6.84 (s, 1H), 6.48 (s, 1H), 4.13 (q, $J = 7.2$ Hz, 2H), 3.86 (s, 3H), 3.82 (s, 3H), 3.78 (s, 3H), 3.84 – 3.73 (m, 3H), 3.67 (m, 1H), 3.35 (t, $J = 5.1$ Hz, 1H), 2.30

(br s, 1H), 1.25 (t, J = 6.9 Hz, 3H), 0.84 (s, 9H), 0.00 (s, 6H); ^{13}C-NMR (CDCl$_3$, 75 MHz): 173.6, 152.1, 148.8, 143.0, 119.8, 114.2, 97.6, 64.8, 62.8, 60.8, 56.7, 56.4, 46.9, 26.0 18.4, 14.5, -5.29, -5.35; ESI MS Calcd for C$_{21}$H$_{37}$NO$_6$Si [M+H] $^+$ 428.25, Found 428.46

4-(2,5-Dimethoxy-phenyl)-but-3-en-1-ol (14). To a stirred solution of 2-Bromo-1,4-dimethoxy-benzene (50.0 mg, 0.23 mmol) and pinacol boronate **13** (78 mg, 0.25 mmol) in toluene (2.9 mL) and methanol (625 µL) was added a solution of K$_2$CO$_3$ (70.0 mg) in water (250 µL). To the resulting suspension, Pd(PPh$_3$)$_4$ (27 mg, 0.023 mmol) was added and the reaction mixture was heated to 60 ºC for 24 hours. The reaction mixture was diluted with ether (5 mL) and filtered through a short pad of Celite. The pad was washed with ether (3 × 5 mL) and the combined layers was concentrated *in vacuo*. The residue was purified by flash chromatography (24:1, hexane:EtOAc) to afford the title compound (**14**) (41 mg, 55 %). ^{1}H-NMR (CDCl$_3$, 300 MHz): δ 6.99 (d, *J* = 2.7 Hz, 1H), 6.83 – 6.73 (m, 3H), 6.19 (dt, *J* = 16.2, 6.9 Hz, 1H), 3.80 (s, 3H), 3.78 (s, 3H), 3.78 – 3.74 (m, 2H), 2.54 – 2.47 (m, 2H); ^{13}C-NMR (CDCl$_3$, 75 MHz): δ 153.9, 151.1, 127.6, 127.5, 127.3, 113.5, 112.3, 112.1, 62.3, 56.4, 56.0, 37.0; ESI MS Calcd for C$_{12}$H$_{16}$O$_3$ [M+H] $^+$ 209.12, Found 209.10

Literature Cited

Literature Cited

(1)   Coombe, D. R.; Kett, W. C. Heparan sulfate-protein interactions: therapeutic
potential through structure-function insights. *Cell Mol Life Sci* **2005***, 62*, 410-424.

(2)   Garg, H. G.; Linhardt, R. J.; Hales, C. A. *Chemistry and Biology of Heparin and
Heparan Sulfate*; Elsevier Science, 2005.

(3)   Raman, R.; Sasisekharan, V.; Sasisekharan, R. Structural insights into biological
roles of protein-glycosaminoglycan interactions. *Chem Biol* **2005***, 12*, 267-277.

(4)   Raghuraman, A.; Riaz, M.; Hindle, M.; Desai, U. R. Rapid and efficient
microwave-assisted synthesis of highly sulfated organic scaffolds. *Tetrahedron
Lett.* **2007***, 48*, 6754-6758.

(5)   Raghuraman, A.; Mosier, P. D.; Desai, U. R. Finding a needle in a haystack:
development of a combinatorial virtual screening approach for identifying high
specificity heparin/heparan sulfate sequence(s). *J Med Chem* **2006***, 49*, 3553-3562.

(6)   McLean, J. The thromboplastic action of cephalin. *Am J Physiol* **1916***, 41*, 250-257.

(7)   Hirsh, J.; Anand, S. S.; Halperin, J. L.; Fuster, V. Guide to anticoagulant therapy:
Heparin: a statement for healthcare professionals from the American Heart
Association. *Circulation* **2001***, 103*, 2994-3018.

(8)   Gallagher, J. T.; Walker, A. Molecular distinctions between heparan sulphate and
heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin

are separate families of N-sulphated polysaccharides. *Biochem J* **1985***, 230*, 665-674.

(9)   Spronk, H. M.; Govers-Riemslag, J. W.; ten Cate, H. The blood coagulation system as a molecular machine. *Bioessays* **2003***, 25*, 1220-1228.

(10)   Murray, C. J.; Lopez, A. D. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **1997***, 349*, 1269-1276.

(11)   Bjork, I.; Olson, S. T. Antithrombin. A bloody important serpin. *Adv Exp Med Biol* **1997***, 425*, 17-33.

(12)   Tollefsen, D. M. Heparin cofactor II. *Adv Exp Med Biol* **1997***, 425*, 35-44.

(13)   Conard, J.; Brosstad, F.; Lie Larsen, M.; Samama, M.; Abildgaard, U. Molar antithrombin concentration in normal human plasma. *Haemostasis* **1983***, 13*, 363-368.

(14)   Tollefsen, D. M.; Pestka, C. A. Heparin cofactor II activity in patients with disseminated intravascular coagulation and hepatic failure. *Blood* **1985***, 66*, 769-774.

(15)   Ishiguro, K.; Kojima, T.; Kadomatsu, K.; Nakayama, Y.; Takagi, A. et al. Complete antithrombin deficiency in mice results in embryonic lethality. *J Clin Invest* **2000***, 106*, 873-878.

(16)   Franco, R. F.; Reitsma, P. H. Genetic risk factors of venous thrombosis. *Hum Genet* **2001***, 109*, 369-384.

(17)   Lane, D. A.; Bayston, T.; Olds, R. J.; Fitches, A. C.; Cooper, D. N. et al. Antithrombin mutation database: 2nd (1997) update. For the Plasma Coagulation

Inhibitors Subcommittee of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost* **1997***, 77*, 197-211.

(18)   Lane, D. A.; Kunz, G.; Olds, R. J.; Thein, S. L. Molecular genetics of antithrombin deficiency. *Blood Rev* **1996***, 10*, 59-74.

(19)   van Boven, H. H.; Lane, D. A. Antithrombin and its inherited deficiency states. *Semin Hematol* **1997***, 34*, 188-204.

(20)   He, L.; Vicente, C. P.; Westrick, R. J.; Eitzman, D. T.; Tollefsen, D. M. Heparin cofactor II inhibits arterial thrombosis after endothelial injury. *J Clin Invest* **2002***, 109*, 213-219.

(21)   Tollefsen, D. M. Heparin cofactor II deficiency. *Arch Pathol Lab Med* **2002***, 126*, 1394-1400.

(22)   Tovar, A. M.; de Mattos, D. A.; Stelling, M. P.; Sarcinelli-Luz, B. S.; Nazareth, R. A. et al. Dermatan sulfate is the predominant antithrombotic glycosaminoglycan in vessel walls: implications for a possible physiological function of heparin cofactor II. *Biochim Biophys Acta* **2005***, 1740*, 45-53.

(23)   Gandossi, E.; Lunven, C.; Gauffeny, C.; Roome, N. O.; Berry, C. N. Platelet aggregation induced in vitro by rabbit plasma clot-associated thrombin, and its inhibition by thrombin inhibitors. *Thromb Haemost* **1998***, 80*, 840-844.

(24)   Huntington, J. A. Mechanisms of glycosaminoglycan activation of the serpins in hemostasis. *J Thromb Haemost* **2003***, 1*, 1535-1549.

(25)   Pike, R. N.; Buckle, A. M.; le Bonniec, B. F.; Church, F. C. Control of the coagulation system by serpins. Getting by with a little help from glycosaminoglycans. *Febs J* **2005**, *272*, 4842-4851.

(26)   van Boeckel, C. A. A.; Petitou, M. The unique antithrombin III binding domain of heparin: A lead to new synthetic antithrombotics. *Angew. Chem. Intl. Ed. Engl.* **1993**, *32*, 1671-1818.

(27)   Olson, S. T.; Bjork, I.; Sheffer, R.; Craig, P. A.; Shore, J. D. et al. Role of the antithrombin-binding pentasaccharide in heparin acceleration of antithrombin-proteinase reactions. Resolution of the antithrombin conformational change contribution to heparin rate enhancement. *J Biol Chem* **1992**, *267*, 12528-12538.

(28)   Olson, S. T.; Bjork, I. Role of protein conformational changes, surface approximation and protein cofactors in heparin-accelerated antithrombin-proteinase reactions. *Adv Exp Med Biol* **1992**, *313*, 155-165.

(29)   Bedsted, T.; Swanson, R.; Chuang, Y. J.; Bock, P. E.; Bjork, I. et al. Heparin and calcium ions dramatically enhance antithrombin reactivity with factor IXa by generating new interaction exosites. *Biochemistry* **2003**, *42*, 8143-8152.

(30)   Rezaie, A. R.; Olson, S. T. Calcium enhances heparin catalysis of the antithrombin-factor Xa reaction by promoting the assembly of an intermediate heparin-antithrombin-factor Xa bridging complex. Demonstration by rapid kinetics studies. *Biochemistry* **2000**, *39*, 12083-12090.

(31) Bray, B.; Lane, D. A.; Freyssinet, J. M.; Pejler, G.; Lindahl, U. Anti-thrombin activities of heparin. Effect of saccharide chain length on thrombin inhibition by heparin cofactor II and by antithrombin. *Biochem J* **1989***, 262*, 225-232.

(32) O'Keeffe, D.; Olson, S. T.; Gasiunas, N.; Gallagher, J.; Baglin, T. P. et al. The heparin binding properties of heparin cofactor II suggest an antithrombin-like activation mechanism. *J Biol Chem* **2004***, 279*, 50267-50273.

(33) Turk, B.; Brieditis, I.; Bock, S. C.; Olson, S. T.; Bjork, I. The oligosaccharide side chain on Asn-135 of alpha-antithrombin, absent in beta-antithrombin, decreases the heparin affinity of the inhibitor by affecting the heparin-induced conformational change. *Biochemistry* **1997***, 36*, 6682-6691.

(34) Baglin, T. P.; Carrell, R. W.; Church, F. C.; Esmon, C. T.; Huntington, J. A. Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism. *Proc Natl Acad Sci U S A* **2002***, 99*, 11079-11084.

(35) Fortenberry, Y. M.; Whinna, H. C.; Gentry, H. R.; Myles, T.; Leung, L. L. et al. Molecular mapping of the thrombin-heparin cofactor II complex. *J Biol Chem* **2004***, 279*, 43237-43244.

(36) Evans, D. L.; McGrogan, M.; Scott, R. W.; Carrell, R. W. Protease specificity and heparin binding and activation of recombinant protease nexin I. *J Biol Chem* **1991***, 266*, 22307-22312.

(37) Pratt, C. W.; Church, F. C. Heparin binding to protein C inhibitor. *J Biol Chem* **1992***, 267*, 8789-8794.

(38)   Pratt, C. W.; Whinna, H. C.; Church, F. C. A comparison of three heparin-binding serine proteinase inhibitors. *J Biol Chem* **1992***, 267*, 8795-8801.

(39)   Rezaie, A. R. Role of exosites 1 and 2 in thrombin reaction with plasminogen activator inhibitor-1 in the absence and presence of cofactors. *Biochemistry* **1999***, 38*, 14592-14599.

(40)   Clowes, A. W.; Karnowsky, M. J. Suppression by heparin of smooth muscle cell proliferation in injured arteries. *Nature* **1977***, 265*, 625-626.

(41)   Burgess, W. H.; Maciag, T. The heparin-binding (fibroblast) growth factor family of proteins. *Annu Rev Biochem* **1989***, 58*, 575-606.

(42)   Shing, Y.; Folkman, J.; Sullivan, R.; Butterfield, C.; Murray, J. et al. Heparin affinity: purification of a tumor-derived capillary endothelial cell growth factor. *Science* **1984***, 223*, 1296-1299.

(43)   Yayon, A.; Klagsbrun, M.; Esko, J. D.; Leder, P.; Ornitz, D. M. Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* **1991***, 64*, 841-848.

(44)   Rapraeger, A. C.; Krufka, A.; Olwin, B. B. Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science* **1991***, 252*, 1705-1708.

(45)   Nugent, M. A.; Edelman, E. R. Kinetics of basic fibroblast growth factor binding to its receptor and heparan sulfate proteoglycan: a mechanism for cooperactivity. *Biochemistry* **1992***, 31*, 8876-8883.

(46)   Ibrahimi, O. A.; Zhang, F.; Hrstka, S. C.; Mohammadi, M.; Linhardt, R. J. Kinetic model for FGF, FGFR, and proteoglycan signal transduction complex assembly. *Biochemistry* **2004***, 43*, 4724-4730.

(47)   Pantoliano, M. W.; Horlick, R. A.; Springer, B. A.; Van Dyk, D. E.; Tobery, T. et al. Multivalent ligand-receptor binding interactions in the fibroblast growth factor system produce a cooperative growth factor and heparin mechanism for receptor dimerization. *Biochemistry* **1994***, 33*, 10229-10248.

(48)   Plotnikov, A. N.; Hubbard, S. R.; Schlessinger, J.; Mohammadi, M. Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell* **2000***, 101*, 413-424.

(49)   Plotnikov, A. N.; Schlessinger, J.; Hubbard, S. R.; Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **1999***, 98*, 641-650.

(50)   Pellegrini, L.; Burke, D. F.; von Delft, F.; Mulloy, B.; Blundell, T. L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **2000***, 407*, 1029-1034.

(51)   Stauber, D. J.; DiGabriele, A. D.; Hendrickson, W. A. Structural interactions of fibroblast growth factor receptor with its ligands. *Proc Natl Acad Sci U S A* **2000***, 97*, 49-54.

(52)   Schlessinger, J.; Plotnikov, A. N.; Ibrahimi, O. A.; Eliseenkova, A. V.; Yeh, B. K. et al. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. *Mol Cell* **2000***, 6*, 743-750.

(53)   Faham, S.; Hileman, R. E.; Fromm, J. R.; Linhardt, R. J.; Rees, D. C. Heparin
         structure and interactions with basic fibroblast growth factor. *Science* **1996**, *271*,
         1116-1120.

(54)   DiGabriele, A. D.; Lax, I.; Chen, D. I.; Svahn, C. M.; Jaye, M. et al. Structure of a
         heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **1998**,
         393, 812-817.

(55)   Gitay-Goren, H.; Soker, S.; Vlodavsky, I.; Neufeld, G. The binding of vascular
         endothelial growth factor to its receptors is dependent on cell surface-associated
         heparin-like molecules. *J Biol Chem* **1992**, *267*, 6093-6098.

(56)   Higashiyama, S.; Abraham, J. A.; Klagsbrun, M. Heparin-binding EGF-like growth
         factor stimulation of smooth muscle cell migration: dependence on interactions
         with cell surface heparan sulfate. *J Cell Biol* **1993**, *122*, 933-940.

(57)   Cohen, T.; Gitay-Goren, H.; Sharon, R.; Shibuya, M.; Halaban, R. et al. VEGF121,
         a vascular endothelial growth factor (VEGF) isoform lacking heparin binding
         ability, requires cell-surface heparan sulfates for efficient binding to the VEGF
         receptors of human melanoma cells. *J Biol Chem* **1995**, *270*, 11322-11326.

(58)   Fannon, M.; Forsten, K. E.; Nugent, M. A. Potentiation and inhibition of bFGF
         binding by heparin: a model for regulation of cellular response. *Biochemistry* **2000**,
         39, 1434-1445.

(59)   Sperinde, G. V.; Nugent, M. A. Heparan sulfate proteoglycans control intracellular
         processing of bFGF in vascular smooth muscle cells. *Biochemistry* **1998**, *37*,
         13153-13164.

(60) Pye, D. A.; Vives, R. R.; Turnbull, J. E.; Hyde, P.; Gallagher, J. T. Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *J Biol Chem* **1998***, 273*, 22936-22942.

(61) Delehedde, M.; Lyon, M.; Gallagher, J. T.; Rudland, P. S.; Fernig, D. G. Fibroblast growth factor-2 binds to small heparin-derived oligosaccharides and stimulates a sustained phosphorylation of p42/44 mitogen-activated protein kinase and proliferation of rat mammary fibroblasts. *Biochem J* **2002***, 366*, 235-244.

(62) Delehedde, M.; Seve, M.; Sergeant, N.; Wartelle, I.; Lyon, M. et al. Fibroblast growth factor-2 stimulation of p42/44MAPK phosphorylation and IkappaB degradation is regulated by heparan sulfate/heparin in rat mammary fibroblasts. *J Biol Chem* **2000***, 275*, 33905-33910.

(63) Walker, A.; Turnbull, J. E.; Gallagher, J. T. Specific heparan sulfate saccharides mediate the activity of basic fibroblast growth factor. *J Biol Chem* **1994***, 269*, 931-935.

(64) Castellot, J. J., Jr.; Wong, K.; Herman, B.; Hoover, R. L.; Albertini, D. F. et al. Binding and internalization of heparin by vascular smooth muscle cells. *J Cell Physiol* **1985***, 124*, 13-20.

(65) Patton, W. A., 2nd; Granzow, C. A.; Getts, L. A.; Thomas, S. C.; Zotter, L. M. et al. Identification of a heparin-binding protein using monoclonal antibodies that block heparin binding to porcine aortic endothelial cells. *Biochem J* **1995***, 311 (Pt 2)*, 461-469.

(66)   Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. Roles of heparan-sulphate glycosaminoglycans in cancer. *Nat Rev Cancer* **2002***, 2*, 521-528.

(67)   Pilia, G.; Hughes-Benzie, R. M.; MacKenzie, A.; Baybayan, P.; Chen, E. Y. et al. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nat Genet* **1996***, 12*, 241-247.

(68)   Lind, T.; Tufaro, F.; McCormick, C.; Lindahl, U.; Lidholt, K. The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J Biol Chem* **1998***, 273*, 26265-26268.

(69)   Sasisekharan, R.; Moses, M. A.; Nugent, M. A.; Cooney, C. L.; Langer, R. Heparinase inhibits neovascularization. *Proc Natl Acad Sci U S A* **1994***, 91*, 1524-1528.

(70)   Sasaki, T.; Larsson, H.; Kreuger, J.; Salmivirta, M.; Claesson-Welsh, L. et al. Structural basis and potential role of heparin/heparan sulfate binding to the angiogenesis inhibitor endostatin. *Embo J* **1999***, 18*, 6240-6248.

(71)   Vlodavsky, I.; Friedmann, Y.; Elkin, M.; Aingorn, H.; Atzmon, R. et al. Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis. *Nat Med* **1999***, 5*, 793-802.

(72)   Vlodavsky, I.; Elkin, M.; Pappo, O.; Aingorn, H.; Atzmon, R. et al. Mammalian heparanase as mediator of tumor metastasis and angiogenesis. *Isr Med Assoc J* **2000***, 2 Suppl*, 37-45.

(73)   Mousa, S. A. Comparative efficacy of different low-molecular-weight heparins (LMWHs) and drug interactions with LMWH: implications for management of vascular disorders. *Semin Thromb Hemost* **2000***, 26 Suppl 1*, 39-46.

(74)   Tyrrell, D. J.; Horne, A. P.; Holme, K. R.; Preuss, J. M.; Page, C. P. Heparin in inflammation: potential therapeutic applications beyond anticoagulation. *Adv Pharmacol* **1999***, 46*, 151-208.

(75)   Tyrell, D. J.; Kilfeather, S.; Page, C. P. Therapeutic uses of heparin beyond its traditional role as an anticoagulant. *Trends Pharmacol Sci* **1995***, 16*, 198-204.

(76)   Nelson, R. M.; Cecconi, O.; Roberts, W. G.; Aruffo, A.; Linhardt, R. J. et al. Heparin oligosaccharides bind L- and P-selectin and inhibit acute inflammation. *Blood* **1993***, 82*, 3253-3258.

(77)   Lantz, M.; Thysell, H.; Nilsson, E.; Olsson, I. On the binding of tumor necrosis factor (TNF) to heparin and the release in vivo of the TNF-binding protein I by heparin. *J Clin Invest* **1991***, 88*, 2026-2031.

(78)   Darien, B. J.; Fareed, J.; Centgraf, K. S.; Hart, A. P.; MacWilliams, P. S. et al. Low molecular weight heparin prevents the pulmonary hemodynamic and pathomorphologic effects of endotoxin in a porcine acute lung injury model. *Shock* **1998***, 9*, 274-281.

(79)   Hiebert, L. M.; Liu, J. M. Heparin protects cultured arterial endothelial cells from damage by toxic oxygen metabolites. *Atherosclerosis* **1990***, 83*, 47-51.

(80) Matzner, Y.; Marx, G.; Drexler, R.; Eldor, A. The inhibitory effect of heparin and related glycosaminoglycans on neutrophil chemotaxis. *Thromb Haemost* **1984**, *52*, 134-137.

(81) Gaffney, P. R.; O'Leary, J. J.; Doyle, C. T.; Gaffney, A.; Hogan, J. et al. Response to heparin in patients with ulcerative colitis. *Lancet* **1991**, *337*, 238-239.

(82) Prajapati, D. N.; Newcomer, J. R.; Emmons, J.; Abu-Hajir, M.; Binion, D. G. Successful treatment of an acute flare of steroid-resistant Crohn's colitis during pregnancy with unfractionated heparin. *Inflamm Bowel Dis* **2002**, *8*, 192-195.

(83) Makrides, S. C. Therapeutic inhibition of the complement system. *Pharmacol Rev* **1998**, *50*, 59-87.

(84) Ecker, E. E.; Gross, P. Anticomplementary power of heparin. *J Infect Dis* **1929**, *44*, 250-253.

(85) Sahu, A.; Pangburn, M. K. Identification of multiple sites of interaction between heparin and the complement system. *Mol Immunol* **1993**, *30*, 679-684.

(86) Maillet, F.; Kazatchkine, M. D.; Glotz, D.; Fischer, E.; Rowe, M. Heparin prevents formation of the human C3 amplification convertase by inhibiting the binding site for B on C3b. *Mol Immunol* **1983**, *20*, 1401-1404.

(87) Cofrancesco, E.; Radaelli, F.; Pogliani, E.; Amici, N.; Torri, G. G. et al. Correlation of sulfate content and degree of carboxylation of heparin and related glycosaminoglycans with anticomplement activity. Relationships to the anticoagulant and platelet-aggregating activities. *Thromb Res* **1979**, *14*, 179-187.

(88)   Raepple, E.; Hill, H. U.; Loos, M. Mode of interaction of different polyanions with the first (C1, C1), the second (C2) and the fourth (C4) component of complement--I. Effect on fluid phase C1 and on C1 bound to EA or to EAC4. *Immunochemistry* **1976***, 13*, 251-255.

(89)   Yu, H.; Munoz, E. M.; Edens, R. E.; Linhardt, R. J. Kinetic studies on the interactions of heparin and complement proteins using surface plasmon resonance. *Biochim Biophys Acta* **2005***, 1726*, 168-176.

(90)   Shukla, D.; Spear, P. G. Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *J Clin Invest* **2001***, 108*, 503-510.

(91)   Barth, H.; Schafer, C.; Adah, M. I.; Zhang, F.; Linhardt, R. J. et al. Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. *J Biol Chem* **2003***, 278*, 41003-41012.

(92)   Argyris, E. G.; Kulkosky, J.; Meyer, M. E.; Xu, Y.; Mukhtar, M. et al. The perlecan heparan sulfate proteoglycan mediates cellular uptake of HIV-1 Tat through a pathway responsible for biological activity. *Virology* **2004***, 330*, 481-486.

(93)   Vives, R. R.; Imberty, A.; Sattentau, Q. J.; Lortat-Jacob, H. Heparan sulfate targets the HIV-1 envelope glycoprotein gp120 coreceptor binding site. *J Biol Chem* **2005***, 280*, 21353-21357.

(94)   WuDunn, D.; Spear, P. G. Initial interaction of herpes simplex virus with cells is binding to heparan sulfate. *J Virol* **1989***, 63*, 52-58.

(95)   Shieh, M. T.; WuDunn, D.; Montgomery, R. I.; Esko, J. D.; Spear, P. G. Cell

surface receptors for herpes simplex virus are heparan sulfate proteoglycans. *J Cell

Biol* **1992***, 116*, 1273-1281.

(96)   Nahmias, A. J.; Kibrick, S. Inhibitory effect of heparin on herpes simplex virus. *J

Bacteriol* **1964***, 87*, 1060-1066.

(97)   Herold, B. C.; WuDunn, D.; Soltys, N.; Spear, P. G. Glycoprotein C of herpes

simplex virus type 1 plays a principal role in the adsorption of virus to cells and in

infectivity. *J Virol* **1991***, 65*, 1090-1098.

(98)   Tal-Singer, R.; Peng, C.; Ponce De Leon, M.; Abrams, W. R.; Banfield, B. W. et al.

Interaction of herpes simplex virus glycoprotein gC with mammalian cell surface

molecules. *J Virol* **1995***, 69*, 4471-4483.

(99)   Williams, R. K.; Straus, S. E. Specificity and affinity of binding of herpes simplex

virus type 2 glycoprotein B to glycosaminoglycans. *J Virol* **1997***, 71*, 1375-1380.

(100)   Shukla, D.; Liu, J.; Blaiklock, P.; Shworak, N. W.; Bai, X. et al. A novel role for 3-

O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* **1999***, 99*, 13-22.

(101)   Liu, J.; Shworak, N. W.; Sinay, P.; Schwartz, J. J.; Zhang, L. et al. Expression of

heparan sulfate D-glucosaminyl 3-O-sulfotransferase isoforms reveals novel

substrate specificities. *J Biol Chem* **1999***, 274*, 5185-5192.

(102)   Trybala, E.; Liljeqvist, J. A.; Svennerholm, B.; Bergstrom, T. Herpes simplex virus

types 1 and 2 differ in their interaction with heparan sulfate. *J Virol* **2000***, 74*,

9106-9114.

(103) Herold, B. C.; Gerber, S. I.; Belval, B. J.; Siston, A. M.; Shulman, N. Differences in the susceptibility of herpes simplex virus types 1 and 2 to modified heparin compounds suggest serotype differences in viral entry. *J Virol* **1996***, 70*, 3461-3469.

(104) Esko, J. D.; Selleck, S. B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu Rev Biochem* **2002***, 71*, 435-471.

(105) Roden, L.; Ananth, S.; Campbell, P.; Curenton, T.; Ekborg, G. et al. *Heparin - an introduction.*; Plenum Press: New York, 1992; 1-20.

(106) Guerrini, M.; Bisio, A.; Torri, G. Combined quantitative (1)H and (13)C nuclear magnetic resonance spectroscopy for characterization of heparin preparations. *Semin Thromb Hemost* **2001***, 27*, 473-482.

(107) Sudo, M.; Sato, K.; Chaidedgumjorn, A.; Toyoda, H.; Toida, T. et al. (1)H nuclear magnetic resonance spectroscopic analysis for determination of glucuronic and iduronic acids in dermatan sulfate, heparin, and heparan sulfate. *Anal Biochem* **2001***, 297*, 42-51.

(108) Maccarana, M.; Sakura, Y.; Tawada, A.; Yoshida, K.; Lindahl, U. Domain structure of heparan sulfates from bovine organs. *J Biol Chem* **1996***, 271*, 17804-17810.

(109) Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M. et al. Conformation of heparin pentasaccharide bound to antithrombin III. *Biochem. J.* **2001***, 359*, 265-272.

(110) Mikhailov, D.; Linhardt, R. J.; Mayo, K. H. NMR solution conformation of heparin-derived hexasaccharide. *Biochem. J.* **1997***, 328*, 51-61.

(111) Mulloy, B.; Forster, M. J. Conformation and dynamics of heparin and heparan sulfate. *Glycobiology* **2000***, 10*, 1147-1156.

(112) Mulloy, B.; Forster, M. J.; Jones, C.; Davies, D. B. NMR and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **1993***, 293*, 849-858.

(113) Mulloy, B.; Forster, M. J.; Jones, C.; Drake, A. F.; Johnson, E. A. et al. The effect of variation of substitution on the solution conformation of heparin: a spectroscopic and molecular modelling study. *Carbohydr. Res.* **1994***, 255*, 1-26.

(114) Ragazzi, M.; Ferro, D. R.; Perly, B.; Sinay, P.; Petitou, M. et al. Conformation of the pentasaccharide corresponding to the binding site of heparin for antithrombin III. *Carbohydr Res* **1990***, 195*, 169-185.

(115) Carter, W. J.; Cama, E.; Huntington, J. A. Crystal structure of thrombin bound to heparin. *J Biol Chem* **2005***, 280*, 2745-2749.

(116) Capila, I.; Hernaiz, M. J.; Mo, Y. D.; Mealy, T. R.; Campos, B. et al. Annexin V--heparin oligosaccharide complex suggests heparan sulfate--mediated assembly on cell surfaces. *Structure* **2001***, 9*, 57-64.

(117) Moon, A. F.; Edavettal, S. C.; Krahn, J. M.; Munoz, E. M.; Negishi, M. et al. Structural analysis of the sulfotransferase (3-o-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1. *J Biol Chem* **2004***, 279*, 45185-45193.

(118) Jin, L.; Abrahams, J. P.; Skinner, R.; Petitou, M.; Pike, R. N. et al. The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci U S A* **1997***, 94*, 14683-14688.

(119) Li, W.; Johnson, D. J.; Esmon, C. T.; Huntington, J. A. Structure of the antithrombin-thrombin-heparin ternary complex reveals the antithrombotic mechanism of heparin. *Nat Struct Mol Biol* **2004***, 11*, 857-862.

(120) Johnson, D. J.; Huntington, J. A. Crystal structure of antithrombin in a heparin-bound intermediate state. *Biochemistry* **2003***, 42*, 8712-8719.

(121) Shaw, J. P.; Johnson, Z.; Borlat, F.; Zwahlen, C.; Kungl, A. et al. The X-ray structure of RANTES: heparin-derived disaccharides allows the rational design of chemokine inhibitors. *Structure* **2004***, 12*, 2081-2093.

(122) Dementiev, A.; Petitou, M.; Herbert, J. M.; Gettins, P. G. The ternary complex of antithrombin-anhydrothrombin-heparin reveals the basis of inhibitor specificity. *Nat Struct Mol Biol* **2004***, 11*, 863-867.

(123) Atkins, E. D.; Nieduszynski, I. A. Crystalline structure of heparin. *Adv Exp Med Biol* **1975***, 52*, 19-37.

(124) Nieduszynski, I. A. *Connective Tissue Polysaccharides.*; VCH: Weinheim, Germany, 1985; 107-140.

(125) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Torri, G.; Casu, B. et al. Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligo-saccharides: NMR and force-field studies. *J Am Chem Soc* **1986***, 108*, 6773-6778.

(126) Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G. et al. Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr Res* **1990***, 195*, 157-167.

(127) Bernfield, M.; Kokenyesi, R.; Kato, M.; Hinkes, M. T.; Spring, J. et al. Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu Rev Cell Biol* **1992***, 8*, 365-393.

(128) Filmus, J.; Selleck, S. B. Glypicans: proteoglycans with a surprise. *J Clin Invest* **2001***, 108*, 497-501.

(129) Iozzo, R. V.; Cohen, I. R.; Grassel, S.; Murdoch, A. D. The biology of perlecan: the multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices. *Biochem J* **1994***, 302 (Pt 3)*, 625-639.

(130) Kato, M.; Wang, H.; Bernfield, M.; Gallagher, J. T.; Turnbull, J. E. Cell surface syndecan-1 on distinct cell types differs in fine structure and ligand binding of its heparan sulfate chains. *J Biol Chem* **1994***, 269*, 18881-18890.

(131) Lyon, M.; Deakin, J. A.; Gallagher, J. T. Liver heparan sulfate structure. A novel molecular design. *J Biol Chem* **1994***, 269*, 11208-11215.

(132) Dietrich, C. P.; Nader, H. B.; Straus, A. H. Structural differences of heparan sulfates according to the tissue and species of origin. *Biochem Biophys Res Commun* **1983***, 111*, 865-871.

(133) Edge, A. S.; Spiro, R. G. Characterization of novel sequences containing 3-O-sulfated glucosamine in glomerular basement membrane heparan sulfate and

localization of sulfated disaccharides to a peripheral domain. *J Biol Chem* **1990**, *265*, 15874-15881.

(134) Tekotte, H.; Engel, M.; Margolis, R. U.; Margolis, R. K. Disaccharide composition of heparan sulfates: brain, nervous tissue storage organelles, kidney, and lung. *J Neurochem* **1994**, *62*, 1126-1130.

(135) Lindahl, U.; Kusche-Gullberg, M.; Kjellen, L. Regulated diversity of heparan sulfate. *J Biol Chem* **1998**, *273*, 24979-24982.

(136) Feyzi, E.; Saldeen, T.; Larsson, E.; Lindahl, U.; Salmivirta, M. Age-dependent modulation of heparan sulfate structure and function. *J Biol Chem* **1998**, *273*, 13395-13398.

(137) Merry, C. L.; Lyon, M.; Deakin, J. A.; Hopwood, J. J.; Gallagher, J. T. Highly sensitive sequencing of the sulfated domains of heparan sulfate. *J Biol Chem* **1999**, *274*, 18455-18462.

(138) Safaiyan, F.; Lindahl, U.; Salmivirta, M. Structural diversity of N-sulfated heparan sulfate domains: distinct modes of glucuronyl C5 epimerization, iduronic acid 2-O-sulfation, and glucosamine 6-O-sulfation. *Biochemistry* **2000**, *39*, 10823-10830.

(139) Jayson, G. C.; Vives, C.; Paraskeva, C.; Schofield, K.; Coutts, J. et al. Coordinated modulation of the fibroblast growth factor dual receptor mechanism during transformation from human colon adenoma to carcinoma. *Int J Cancer* **1999**, *82*, 298-304.

(140) Brickman, Y. G.; Ford, M. D.; Gallagher, J. T.; Nurcombe, V.; Bartlett, P. F. et al. Structural modification of fibroblast growth factor-binding heparan sulfate at a determinative stage of neural development. *J Biol Chem* **1998***, 273*, 4350-4359.

(141) Dhoot, G. K.; Gustafsson, M. K.; Ai, X.; Sun, W.; Standiford, D. M. et al. Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. *Science* **2001***, 293*, 1663-1666.

(142) Ai, X.; Do, A. T.; Lozynska, O.; Kusche-Gullberg, M.; Lindahl, U. et al. QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. *J Cell Biol* **2003***, 162*, 341-351.

(143) Folkman, J.; Klagsbrun, M.; Sasse, J.; Wadzinski, M.; Ingber, D. et al. A heparin-binding angiogenic protein--basic fibroblast growth factor--is stored within basement membrane. *Am J Pathol* **1988***, 130*, 393-400.

(144) Vlodavsky, I.; Folkman, J.; Sullivan, R.; Fridman, R.; Ishai-Michaeli, R. et al. Endothelial cell-derived basic fibroblast growth factor: synthesis and deposition into subendothelial extracellular matrix. *Proc Natl Acad Sci U S A* **1987***, 84*, 2292-2296.

(145) Perrimon, N.; Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **2000***, 404*, 725-728.

(146) Atha, D. H.; Stephens, A. W.; Rosenberg, R. D. Evaluation of critical groups required for the binding of heparin to antithrombin. *Proc Natl Acad Sci U S A* **1984***, 81*, 1030-1034.

(147) Lindahl, U.; Backstrom, G.; Thunberg, L.; Leder, I. G. Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin. *Proc Natl Acad Sci U S A* **1980**, *77*, 6551-6555.

(148) Liu, J.; Shriver, Z.; Pope, R. M.; Thorp, S. C.; Duncan, M. B. et al. Characterization of a heparan sulfate octasaccharide that binds to herpes simplex virus type 1 glycoprotein D. *J Biol Chem* **2002**, *277*, 33456-33467.

(149) Turnbull, J. E.; Fernig, D. G.; Ke, Y.; Wilkinson, M. C.; Gallagher, J. T. Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate. *J Biol Chem* **1992**, *267*, 10337-10341.

(150) Guimond, S.; Maccarana, M.; Olwin, B. B.; Lindahl, U.; Rapraeger, A. C. Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4. *J Biol Chem* **1993**, *268*, 23906-23914.

(151) Maccarana, M.; Casu, B.; Lindahl, U. Minimal sequence in heparin/heparan sulfate required for binding of basic fibroblast growth factor. *J Biol Chem* **1993**, *268*, 23898-23905.

(152) Lundin, L.; Larsson, H.; Kreuger, J.; Kanda, S.; Lindahl, U. et al. Selectively desulfated heparin inhibits fibroblast growth factor-induced mitogenicity and angiogenesis. *J Biol Chem* **2000**, *275*, 24653-24660.

(153) Gama, C. I.; Tully, S. E.; Sotogaku, N.; Clark, P. M.; Rawat, M. et al. Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat Chem Biol* **2006**, *2*, 467-473.

(154) Venkataraman, G.; Shriver, Z.; Raman, R.; Sasisekharan, R. Sequencing complex polysaccharides. *Science* **1999***, 286*, 537-542.

(155) Keiser, N.; Venkataraman, G.; Shriver, Z.; Sasisekharan, R. Direct isolation and sequencing of specific protein-binding glycosaminoglycans. *Nat Med* **2001***, 7*, 123-128.

(156) Noti, C.; Seeberger, P. H. Chemical approaches to define the structure-activity relationship of heparin-like glycosaminoglycans. *Chem Biol* **2005***, 12*, 731-756.

(157) Seeberger, P. H. Automated oligosaccharide synthesis. *Chem Soc Rev* **2008***, 37*, 19-28.

(158) Polat, T.; Wong, C. H. Anomeric reactivity-based one-pot synthesis of heparin-like oligosaccharides. *J Am Chem Soc* **2007***, 129*, 12795-12800.

(159) Chen, J.; Jones, C. L.; Liu, J. Using an enzymatic combinatorial approach to identify anticoagulant heparan sulfate structures. *Chem Biol* **2007***, 14*, 986-993.

(160) Linhardt, R. J.; Kim, J. H. Combinatorial enzymatic synthesis of heparan sulfate. *Chem Biol* **2007***, 14*, 972-973.

(161) Jemth, P.; Kreuger, J.; Kusche-Gullberg, M.; Sturiale, L.; Gimenez-Gallego, G. et al. Biosynthetic oligosaccharide libraries for identification of protein-binding heparan sulfate motifs. Exploring the structural diversity by screening for fibroblast growth factor (FGF)1 and FGF2 binding. *J Biol Chem* **2002***, 277*, 30567-30573.

(162) Feizi, T.; Fazio, F.; Chai, W.; Wong, C. H. Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics. *Curr Opin Struct Biol* **2003***, 13*, 637-645.

(163)   Wang, D. Carbohydrate microarrays. *Proteomics* **2003***, 3*, 2167-2175.

(164)   Tully, S. E.; Rawat, M.; Hsieh-Wilson, L. C. Discovery of a TNF-alpha antagonist using chondroitin sulfate microarrays. *J Am Chem Soc* **2006***, 128*, 7740-7741.

(165)   de Paz, J. L.; Noti, C.; Seeberger, P. H. Microarrays of synthetic heparin oligosaccharides. *J Am Chem Soc* **2006***, 128*, 2766-2767.

(166)   Shipp, E. L.; Hsieh-Wilson, L. C. Profiling the sulfation specificities of glycosaminoglycan interactions with growth factors and chemotactic proteins using microarrays. *Chem Biol* **2007***, 14*, 195-208.

(167)   Desai, U. R. New antithrombin-based anticoagulants. *Med Res Rev* **2004***, 24*, 151-181.

(168)   Desai, U. R.; Petitou, M.; Bjork, I.; Olson, S. T. Mechanism of heparin activation of antithrombin. Role of individual residues of the pentasaccharide activating sequence in the recognition of native and activated states of antithrombin. *J Biol Chem* **1998***, 273*, 7478-7487.

(169)   Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J Comput Aided Mol Des* **1991***, 5*, 545-552.

(170)   Gunnarsson, G. T.; Desai, U. R. Designing small, nonsugar activators of antithrombin using hydropathic interaction analyses. *J Med Chem* **2002***, 45*, 1233-1243.

(171)   Gunnarsson, G. T.; Desai, U. R. Interaction of designed sulfated flavanoids with antithrombin: lessons on the design of organic activators. *J Med Chem* **2002***, 45*, 4460-4470.

(172)   Dantuluri, M.; Gunnarsson, G. T.; Riaz, M.; Nguyen, H.; Desai, U. R. Capillary electrophoresis of highly sulfated flavanoids and flavonoids. *Anal Biochem* **2005***, 336*, 316-322.

(173)   Gunnarsson, G. T.; Riaz, M.; Adams, J.; Desai, U. R. Synthesis of per-sulfated flavonoids using 2,2,2-trichloro ethyl protecting group and their factor Xa inhibition potential. *Bioorg Med Chem* **2005***, 13*, 1783-1789.

(174)   Bogenstätter, M.; Limberg, A.; Overman, L. E.; Tomasi, A. L. Enantioselective Total Synthesis of the Kinesin Motor Protein Inhibitor Adociasulfate 1. *J Am Chem Soc* **1999***, 121*, 12206-12207.

(175)   Simpson, L. S.; Widlanski, T. S. A comprehensive approach to the synthesis of sulfate esters. *J Am Chem Soc* **2006***, 128*, 1605-1610.

(176)   Santos, G. A.; Murray, A. P.; Pujol, C. A.; Damonte, E. B.; Maier, M. S. Synthesis and antiviral activity of sulfated and acetylated derivatives of 2beta,3alpha-dihydroxy-5alpha-cholestane. *Steroids* **2003***, 68*, 125-132.

(177)   Kawai, N.; Takao, K.; Kobayashi, S. Synthetic study of akaterpin: Determination of the relative stereochemistry of the upper decalin moiety with disulfated hydroquin. *Tetrahedron Lett.* **1999***, 40*, 4193-4196.

(178)   Liu, Y.; Lien, I. F.; Ruttgaizer, S.; Dove, P.; Taylor, S. D. Synthesis and protection of aryl sulfates using the 2,2,2-trichloroethyl moiety. *Org Lett* **2004***, 6*, 209-212.

(179)   Lu, L. D.; Shie, C. R.; Kulkarni, S. S.; Pan, G. R.; Lu, X. A. et al. Synthesis of 48 disaccharide building blocks for the assembly of a heparin and heparan sulfate oligosaccharide library. *Org Lett* **2006***, 8*, 5995-5998.

(180)   Fan, R. H.; Achkar, J.; Hernandez-Torres, J. M.; Wei, A. Orthogonal sulfation strategy for synthetic heparan sulfate ligands. *Org Lett* **2005***, 7*, 5095-5098.

(181)   Orgueira, H. A.; Bartolozzi, A.; Schell, P.; Litjens, R. E.; Palmacci, E. R. et al. Modular synthesis of heparin oligosaccharides. *Chemistry* **2003***, 9*, 140-169.

(182)   Codée, J. C. C.; van der Marel, G. A.; van Boeckel, C. A. A.; van Boom, J. H. Probing the Heparin-Antithrombin III Interaction Using Synthetic Pentasaccharides Bearing Positively Charged Groups. *Eur. J. Org. Chem.* **2002**, 3954-3965.

(183)   Kappe, C. O.; Dallinger, D. The impact of microwave synthesis on drug discovery. *Nat Rev Drug Discov* **2006***, 5*, 51-63.

(184)   Gabriel, C.; Gabriel, S.; Grant, E. H.; Grant, E. H.; Halstead, B. S. et al. Dielectric parameters relevant to microwave dielectric heating. *Chem. Soc. Rev.* **1998***, 27*, 213-224.

(185)   Petitou, M.; van Boeckel, C. A. A synthetic antithrombin III binding pentasaccharide is now a drug! What comes next? *Angew Chem Int Ed Engl* **2004***, 43*, 3118-3133.

(186)   Bates, S. M.; Weitz, J. I. New anticoagulants: beyond heparin, low-molecular-weight heparin and warfarin. *Br J Pharmacol* **2005***, 144*, 1017-1028.

(187)   Bates, S. M.; Weitz, J. I. The status of new anticoagulants. *Br J Haematol* **2006***, 134*, 3-19.

(188) Skinner, R.; Abrahams, J. P.; Whisstock, J. C.; Lesk, A. M.; Carrell, R. W. et al. The 2.6 A structure of antithrombin indicates a conformational change at the heparin binding site. *J Mol Biol* **1997***, 266*, 601-609.

(189) Izaguirre, G.; Zhang, W.; Swanson, R.; Bedsted, T.; Olson, S. T. Localization of an antithrombin exosite that promotes rapid inhibition of factors Xa and IXa dependent on heparin activation of the serpin. *J Biol Chem* **2003***, 278*, 51433-51440.

(190) Arocas, V.; Bock, S. C.; Raja, S.; Olson, S. T.; Bjork, I. Lysine 114 of antithrombin is of crucial importance for the affinity and kinetics of heparin pentasaccharide binding. *J Biol Chem* **2001***, 276*, 43809-43817.

(191) Desai, U.; Swanson, R.; Bock, S. C.; Bjork, I.; Olson, S. T. Role of arginine 129 in heparin binding and activation of antithrombin. *J Biol Chem* **2000***, 275*, 18976-18984.

(192) Schedin-Weiss, S.; Desai, U. R.; Bock, S. C.; Gettins, P. G.; Olson, S. T. et al. Importance of lysine 125 for heparin binding and activation of antithrombin. *Biochemistry* **2002***, 41*, 4779-4788.

(193) Monien, B. H.; Krishnasamy, C.; Olson, S. T.; Desai, U. R. Importance of tryptophan 49 of antithrombin in heparin binding and conformational activation. *Biochemistry* **2005***, 44*, 11660-11668.

(194) Schedin-Weiss, S.; Desai, U. R.; Bock, S. C.; Olson, S. T.; Bjork, I. Roles of N-terminal region residues Lys11, Arg13, and Arg24 of antithrombin in heparin

recognition and in promotion and stabilization of the heparin-induced conformational change. *Biochemistry* **2004***, 43*, 675-683.

(195) Arocas, V.; Bock, S. C.; Olson, S. T.; Bjork, I. The role of Arg46 and Arg47 of antithrombin in heparin binding. *Biochemistry* **1999***, 38*, 10196-10204.

(196) Arocas, V.; Turk, B.; Bock, S. C.; Olson, S. T.; Bjork, I. The region of antithrombin interacting with full-length heparin chains outside the high-affinity pentasaccharide sequence extends to Lys136 but not to Lys139. *Biochemistry* **2000***, 39*, 8512-8518.

(197) Gassman, P. G.; Schenk, W. N. A general procedure for the base-promoted hydrolysis of hindered esters at ambient temperatures. *J. Org. Chem.* **1977***, 42*, 918-920.

(198) Schedin-Weiss, S.; Arocas, V.; Bock, S. C.; Olson, S. T.; Bjork, I. Specificity of the basic side chains of Lys114, Lys125, and Arg129 of antithrombin in heparin binding. *Biochemistry* **2002***, 41*, 12369-12376.

(199) Rabenstein, D. L. Heparin and heparan sulfate: structure and function. *Nat Prod Rep* **2002***, 19*, 312-331.

(200) Cardin, A. D.; Weintraub, H. J. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **1989***, 9*, 21-32.

(201) Margalit, H.; Fischer, N.; Ben-Sasson, S. A. Comparative analysis of structurally defined heparin binding sequences reveals a distinct spatial distribution of basic residues. *J Biol Chem* **1993***, 268*, 19228-19231.

(202)   Hileman, R. E.; Fromm, J. R.; Weiler, J. M.; Linhardt, R. J. Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* **1998**, *20*, 156-167.

(203)   Thompson, L. D.; Pantoliano, M. W.; Springer, B. A. Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry* **1994**, *33*, 3831-3840.

(204)   Olson, S. T.; Halvorson, H. R.; Bjork, I. Quantitative characterization of the thrombin-heparin interaction. Discrimination between specific and nonspecific binding models. *J Biol Chem* **1991**, *266*, 6342-6352.

(205)   Bitomsky, W. W., R.C. Docking of Glycosaminoglycans to Heparin-Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL-8. *J Am Chem Soc* **1999**, *121*, 3004-3013.

(206)   Boeckel, P. D. J. G. a. C. A. A. V. Constructing a molecular model of the interaction between antithrombin III and a potent heparin analog. *J Am Chem Soc* **1991**, *113*, 2743.

(207)   Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology* **1997**, *267*, 727-748.

(208)   Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609-623.

(209) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem* **2004***, 47*, 45-55.

(210) Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y. et al. Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J Chem Inf Comput Sci* **2004***, 44*, 871-881.

(211) Lindahl, U.; Thunberg, L.; Backstrom, G.; Riesenfeld, J.; Nordling, K. et al. Extension and structural variability of the antithrombin-binding sequence in heparin. *J Biol Chem* **1984***, 259*, 12368-12376.

(212) Belzar, K. J.; Dafforn, T. R.; Petitou, M.; Carrell, R. W.; Huntington, J. A. The effect of a reducing-end extension on pentasaccharide binding by antithrombin. *J Biol Chem* **2000***, 275*, 8733-8741.

(213) Desai, U. R.; Petitou, M.; Bjork, I.; Olson, S. T. Mechanism of heparin activation of antithrombin: evidence for an induced-fit model of allosteric activation involving two interaction subsites. *Biochemistry* **1998***, 37*, 13033-13041.

(214) Grootenhuis, P. D.; Van Boeckel, C. A. Constructing a molecular model of the interaction between antithrombin III and a potent heparin analog. *J Am Chem Soc* **1991***, 113*, 2743 - 2747.

(215) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004***, 3*, 935-949.

(216) Gettins, P. G. Serpin structure, mechanism, and function. *Chem Rev* **2002**, *102*, 4751-4804.

(217) Maimone, M. M.; Tollefsen, D. M. Activation of heparin cofactor II by heparin oligosaccharides. *Biochem Biophys Res Commun* **1988**, *152*, 1056-1061.

(218) Olson, S. T.; Bjork, I.; Shore, J. D. Kinetic characterization of heparin-catalyzed and uncatalyzed inhibition of blood coagulation proteinases by antithrombin. *Methods Enzymol* **1993**, *222*, 525-559.

(219) Aihara, K.; Azuma, H.; Takamori, N.; Kanagawa, Y.; Akaike, M. et al. Heparin cofactor II is a novel protective factor against carotid atherosclerosis in elderly individuals. *Circulation* **2004**, *109*, 2761-2765.

(220) Takamori, N.; Azuma, H.; Kato, M.; Hashizume, S.; Aihara, K. et al. High plasma heparin cofactor II activity is associated with reduced incidence of in-stent restenosis after percutaneous coronary intervention. *Circulation* **2004**, *109*, 481-486.

(221) Bendayan, P.; Boccalon, H.; Dupouy, D.; Boneu, B. Dermatan sulfate is a more potent inhibitor of clot-bound thrombin than unfractionated and low molecular weight heparins. *Thromb Haemost* **1994**, *71*, 576-580.

(222) Chuang, Y. J.; Swanson, R.; Raja, S. M.; Olson, S. T. Heparin enhances the specificity of antithrombin for thrombin and factor Xa independent of the reactive center loop sequence. Evidence for an exosite determinant of factor Xa specificity in heparin-activated antithrombin. *J Biol Chem* **2001**, *276*, 14961-14971.

(223)   Izaguirre, G.; Olson, S. T. Residues Tyr253 and Glu255 in strand 3 of beta-sheet C of antithrombin are key determinants of an exosite made accessible by heparin activation to promote rapid inhibition of factors Xa and IXa. *J Biol Chem* **2006**, *281*, 13424-13432.

(224)   Liaw, P. C.; Becker, D. L.; Stafford, A. R.; Fredenburgh, J. C.; Weitz, J. I. Molecular basis for the susceptibility of fibrin-bound thrombin to inactivation by heparin cofactor ii in the presence of dermatan sulfate but not heparin. *J Biol Chem* **2001**, *276*, 20959-20965.

(225)   Verhamme, I. M.; Bock, P. E.; Jackson, C. M. The preferred pathway of glycosaminoglycan-accelerated inactivation of thrombin by heparin cofactor II. *J Biol Chem* **2004**, *279*, 9785-9795.

(226)   Warda, M.; Gouda, E. M.; Toida, T.; Chi, L.; Linhardt, R. J. Isolation and characterization of raw heparin from dromedary intestine: evaluation of a new source of pharmaceutical heparin. *Comp Biochem Physiol C Toxicol Pharmacol* **2003**, *136*, 357-365.

(227)   Warda, M.; Linhardt, R. J. Dromedary glycosaminoglycans: molecular characterization of camel lung and liver heparan sulfate. *Comp Biochem Physiol B Biochem Mol Biol* **2006**, *143*, 37-43.

(228)   Vongchan, P.; Warda, M.; Toyoda, H.; Toida, T.; Marks, R. M. et al. Structural characterization of human liver heparan sulfate. *Biochim Biophys Acta* **2005**, *1721*, 1-8.

(229) Blinder, M. A.; Tollefsen, D. M. Site-directed mutagenesis of arginine 103 and lysine 185 in the proposed glycosaminoglycan-binding site of heparin cofactor II. *J Biol Chem* **1990***, 265*, 286-291.

(230) Liaw, P. C.; Austin, R. C.; Fredenburgh, J. C.; Stafford, A. R.; Weitz, J. I. Comparison of heparin- and dermatan sulfate-mediated catalysis of thrombin inactivation by heparin cofactor II. *J Biol Chem* **1999***, 274*, 27597-27604.

(231) Whinna, H. C.; Blinder, M. A.; Szewczyk, M.; Tollefsen, D. M.; Church, F. C. Role of lysine 173 in heparin binding to heparin cofactor II. *J Biol Chem* **1991***, 266*, 8129-8135.

(232) Skinner, R.; Chang, W. S.; Jin, L.; Pei, X.; Huntington, J. A. et al. Implications for function and therapy of a 2.9 A structure of binary-complexed antithrombin. *J Mol Biol* **1998***, 283*, 9-14.

(233) Fitton, H. L.; Skinner, R.; Dafforn, T. R.; Jin, L.; Pike, R. N. The N-terminal segment of antithrombin acts as a steric gate for the binding of heparin. *Protein Sci* **1998***, 7*, 782-788.

(234) Zhang, Y.; McElrea, A.; Sanchez, G. V., Jr.; Do, D.; Gomez, A. et al. Dicationic electrophiles from olefinic amines in superacid. *J Org Chem* **2003***, 68*, 5119-5122.

(235) Solorio, D. M.; Jennings, M. P. Total synthesis and absolute configuration determination of (+)-bruguierol C. *J Org Chem* **2007***, 72*, 6621-6623.

(236) Corey, E. J.; Ensley, H. E.; Suggs, J. W. Convenient synthesis of (S)-(-)-pulegone from (-)-citronellol. *J Org Chem* **1976***, 41*, 380 - 381.

(237) Li, C. F.; Liu, H.; Liao, J.; Cao, Y. J.; Liu, X. P. et al. Enantioselective organocatalytic intramolecular ring-closing Friedel-Crafts-type alkylation of indoles. *Org Lett* **2007**, *9*, 1847-1850.

(238) Bartoli, G.; Bosco, M.; Carlone, A.; Pesciaioli, F.; Sambri, L. et al. Organocatalytic asymmetric Friedel-Crafts alkylation of indoles with simple alpha,beta-unsaturated ketones. *Org Lett* **2007**, *9*, 1403-1405.

(239) Katagiri, N.; Kato, T.; Nakano, J. Studies on ketene and its derivatives. CIX. Synthesis of naturally occurring anthracene-9,10-diones. *Chem Pharm Bull* **1982**, *30*, 2440-2446.

(240) Nakano, J.; Katagiri, N.; Kato, T. Studies on ketene and its derivatives. CX. Synthesis of 1,3-dimethoxyfluoren-9-ones. *Chem Pharm Bull* **1982**, *30*, 2590-2594.

(241) Chrzanowska, M.; Rozwadowska, M. D. Asymmetric synthesis of isoquinoline alkaloids. *Chem Rev* **2004**, *104*, 3341-3370.

(242) Shigehisa, H.; Takayama, J.; T., H. The first total synthesis of (±)-annosqualine by means of oxidative enamide–phenol coupling: pronounced effect of phenoxide formation on the phenol oxidation mechanism. *Tetrahedron Lett.* **2006**, *47*, 7301-7306.

(243) Zhu, J.; Lu, J.; Zhou, Y.; Li, Y.; Cheng, J. et al. Design, synthesis, and antifungal activities in vitro of novel tetrahydroisoquinoline compounds based on the structure of lanosterol 14α-demethylase (CYP51) of fungi. *Bioorg Med Chem Lett* **2006**, *16*, 5285-5289.

(244)   Ortiz, J. C.; Ozores, L.; Cagide-Fagín, F.; Alonso, R. Annulation of -aryl--nitro-,-enals and 2,2-dimethyl-1,3-dioxan-5-one: a one-step assembly of nitrocyclitols. Application to a short practical synthesis of (±)-7-deoxy-2-epi-pancratistatin tetraacetate. *Chem. Commun.* **2006**, 4239-4241.

APPENDIX A

A1. Experimental procedures and spectral data for synthetic schemes leading to IES5 and IAS5 [*]

Figure 50. Synthesis of **IES5,** an analog of designed antithrombin activator **IAS5**
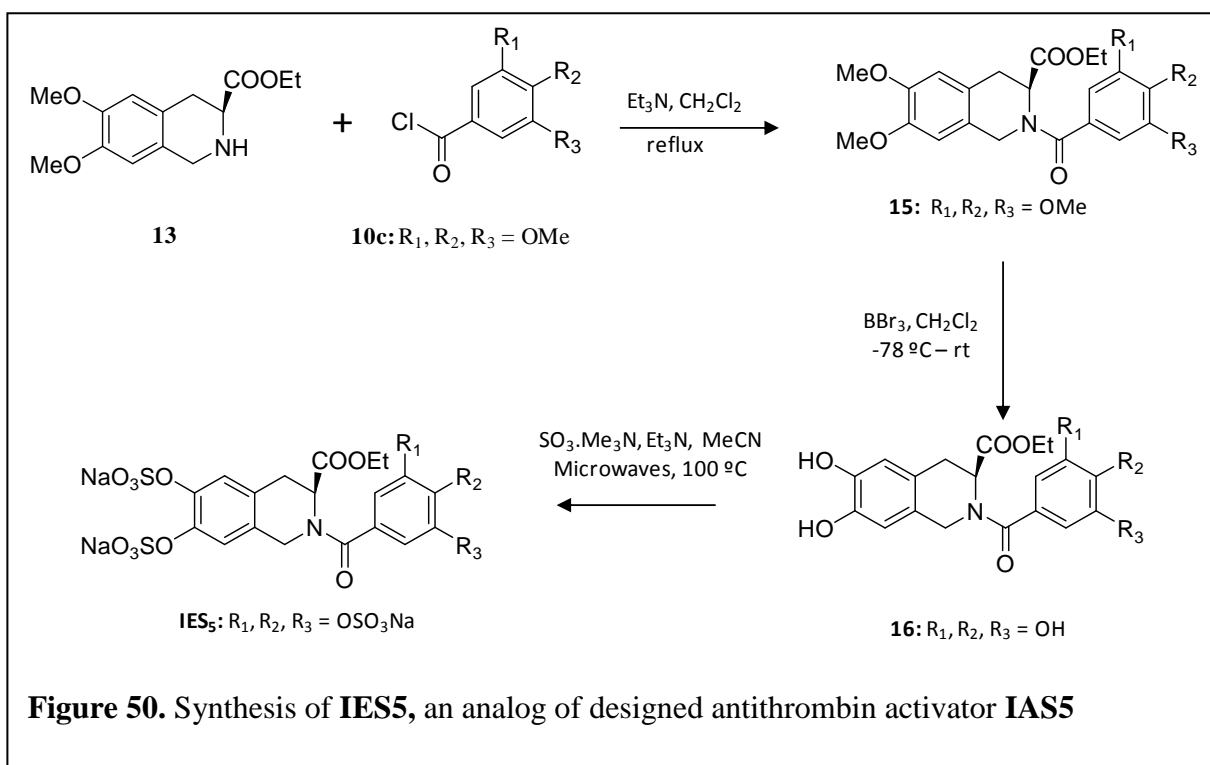
Synthesis of amide **15**: To a stirred suspension of amine 13 (2.25 g, 8.5 mmol),

triethylamine (5.9 mL, 42.5 mmol) in dichloromethane (43 mL) at 0 ºC, was added acid

---

[*] Spectral data for **IES4** and intermediates leading to it may be found in Chapter 3

chloride **10c** (2.05 g, 8.9 mmol) in portions. The reaction mixture was allowed to warm to room temperature and refluxed. After 3-4 hrs, the reaction mixture diluted with dichloromethane (50 mL), washed with 0.5 N HCl (3×50 mL) and 5 % potassium carbonate (3×50 mL), dried over anhydrous Na_2SO_4, and concentrated *in vacuo* to obtain a colorless oil (3.2 g, 82 %). **15:** ^{1}H NMR (400 MHz, CDCl$_3$): δ 6.26-6.49 (m, 4H), 5.21 (br s, 1H, isomer I), 4.83 (d, *J* = 17.2 Hz, 1H, isomer 1), 4.60 (br s, 1 H, isomer II), 4.30-4.42 (m, 2H, isomer II), 4.23 (d, *J* = 17.2 Hz, 1H, isomer 1), 3.78-3.91 (m, 2H, isomers I-II), 3.49-3.60 (m, 15H), 2.88-2.96 (m, 2H), 0.85-0.98 (m, 3H, isomers I-II)

Polyphenols **16**: To a stirred solution of the amide (3.2 g, 7.0 mmol) in dichloromethane (80 mL) at -78 ºC, was added BBr$_3$ (42 mL of 1M solution in CH$_2$Cl$_2$, 1.2 equiv per OMe group) under N$_2$ over 15 minutes. After stirring for 12 hrs at rt, the reaction was quenched at 0 °C with MeOH (10 mL) and water (10 mL). The reaction mixture was partitioned between EtOAc (220 mL) and 2N HCl (50 mL). The aqueous layer was diluted with brine (50 mL) and washed with EtOAc (6 x 50 mL). The combined organic layer was dried over Na$_2$SO$_4$ and concentrated *in vacuo*. The residue was purified by silica gel chromatography (Hexanes/EtOAc = 1:1, 1:4, 1:4.5, 0:1) to give a yellow solid (1.9 g, 68%). **16:** ^{1}H NMR (400 MHz, DMSO-*d6*): δ 9.18 (s, -OH), 9.16 (s -OH), 8.82 (m, -OH), 8.52 (m, -OH), 6.29-6.53 (m, 4H), 5.01 (m, 1H, isomer I), 4.84 (m, 1H, isomer I), 4.70 (m, 1H, isomer II), 4.36 (m, 2H, isomer II), 4.19 (m, 1H, isomer I), 3.55 (m, 2H, isomers I-II), 2.95, (m, 2H, isomers I-II), 1.15 (m, 3H, isomers I-II);

Per-sulfate **IES$_5$**: To a stirred solution of the poly-alcohol (20 mg, 0.05 mmol) in

MeCN (1 mL) at rt, Et$_3$N (0.35 mL, 2.5 mmol) and Me$_3$N.SO$_3$ (313 mg, 2.3 mmol) was added. The reaction vessel was sealed and micro-waved for 20 minutes at 100 ºC. The reaction was repeated for 4 times and the reaction mixture was pooled together. The MeCN layer was decanted and pooled, while the residue from each tube was washed with MeCN (5 mL) and centrifuged. The combined MeCN layers were concentrated *in vacuo*. Water (5 mL) was added to the residue and stirred for 10 min. The water layer was concentrated to approximately 2 mL, loaded onto a Sephadex G10 column (~ 160 cm) and chromatographed using water as eluent. Fractions were combined based on RP-HPLC profiles, concentrated and re-loaded onto a SP Sephadex C25 column for sodium exchange. Appropriate fractions were pooled, concentrated *in vacuo* and lyophilized to obtain a white powder (140 mg, 60 %). **IES$_5$:** ^{1}H NMR (DMSO, 400 MHz) δ: 7.30-7.40 (m, 4H, isomers I-II), 5.22 (s, 1H, isomer I), 4.89 (d, *J* = 17.2 Hz, 1H, isomer I), 4.84 (m, 1H, isomer II), 4.38 (m, 2H, isomer II), 4.20 (d, *J* = 17.2 Hz, 1H, isomer I), 3.56-3.61 (m, 2H, isomers I-II), 3.00-3.14 (m, 2H, isomers I-II).

A2. Capillary electropherograms for IAS5 and its analogs

The following electropherograms were recorded to assess the purity of synthesized poly-sulfated compounds **1s**-**8s**. Experimental conditions may be found in section 3.4.1).
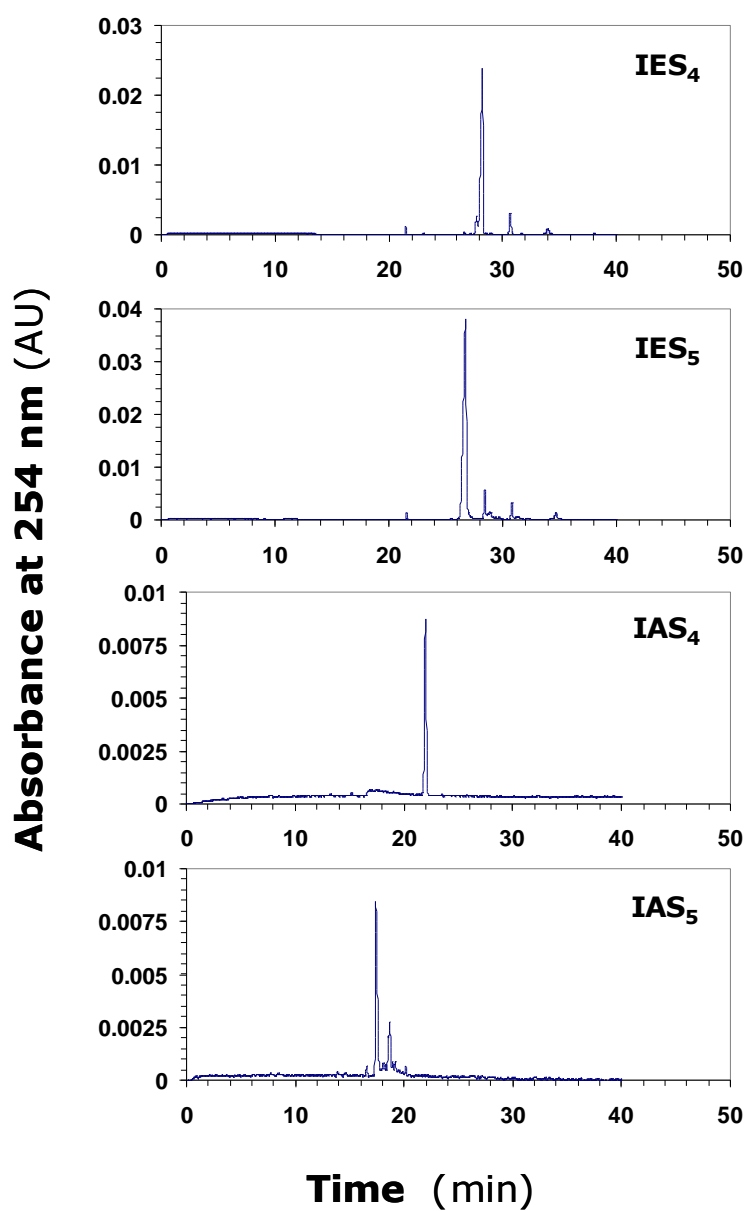
<u>VITA</u>

Arjun Raghuraman was born on 13 November, 1979 in Chennai, India and is an Indian citizen. He obtained is Bachelor's degree in Pharmacy from the SRM Institute of Science and Technology in Chennai in 2001 and worked as an Analytical Chemist at Orchid Health Care, Chennai between 2002 and 2003. He began graduate studies in Department of Medicinal Chemistry at Virginia Commonwealth University, in Richmond, USA in August 2003.

.